MATHEMATICS OF OPERATIONS RESEARCH
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0364-765X | EISSN 1526-5471 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000

Strong Metric (Sub)regularity of KKT Mappings for Piecewise Linear-Quadratic Convex-Composite Optimization and the Quadratic Convergence of Newton's Method

J. V. Burke
University of Washington, Department of Mathematics, Seattle, WA;
jvburke@uw.edu, https://sites.math.washington.edu/~burke

A. Engle
University of Washington, Department of Mathematics, Seattle, WA;
aengle2@uw.edu, https://aengle1429.github.io

This work concerns the local convergence theory of Newton and quasi-Newton methods for *convex-composite* optimization: minimize $f(x) := h(c(x))$, where h is an infinite-valued proper convex function and c is $\mathcal{C}^2$-smooth. We focus on the case where h is infinite-valued piecewise linear-quadratic and convex. Such problems include nonlinear programming, mini-max optimization, estimation of nonlinear dynamics with non-Gaussian noise as well as many modern approaches to large-scale data analysis and machine learning. Our approach embeds the optimality conditions for convex-composite optimization problems into a generalized equation. We establish conditions for strong metric subregularity and strong metric regularity of the corresponding set-valued mappings. This allows us to extend classical convergence of Newton and quasi-Newton methods to the broader class of non-finite valued piecewise linear-*quadratic* convex-composite optimization problems. In particular we establish local quadratic convergence of the Newton method under conditions that parallel those in nonlinear programming when h is non-finite valued piecewise linear.

Key words: Convex-composite optimization, generalized equations, Newton's method, quasi-Newton methods, partial smoothness and active manifold identification, piecewise linear-quadratic, strong metric subregularity, strong metric regularity, quadratic convergence
MSC2000 subject classification: 90C25, 90C26, 90C30, 90C31, 90C33, 90C46, 90C53, 90C55, 65K10

1. Introduction This work concerns local convergence theory of Newton and quasi-Newton methods for the solution of the *convex-composite* problem:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) := h(c(x)), \tag{$\mathbf{P}$}$$

where $h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is piecewise linear-quadratic (PLQ) and convex, and $c : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^2$-smooth. When $h = \frac{1}{2}\|\cdot\|^2$, **P** is the classical nonlinear least-squares problem. Numerous other problems fall within this class including nonlinear programming (NLP), mini-max optimization, estimation of nonlinear dynamics with non-Gaussian noise as well as many modern approaches to large-scale data analysis and machine learning [1, 2, 11]. Convex-composite optimization has a long history with investigations in the 1970s [29, 30], 1980s [3, 4, 22, 34, 35, 39, 40], and 1990s [6, 7, 12, 37], where much of the emphasis was on a calculus for compositions and its relationship to nonlinear programming (NLP) and exact penalization [19]. Recently, there has been a resurgence of interest in local [15, 18] and global [9, 10, 15, 16, 17, 24] algorithms for this class of problems especially with respect to establishing the iteration complexity of first-order methods for **P**. Much of this work has focused on the case where the function h is finite-valued.

These, as well as most methods for solving **P**, use a direction-finding subproblem similar to

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad h(c(x^k)+\nabla c(x^k)[x-x^k])+\frac{1}{2}[x-x^k]^\top H_k[x-x^k], \tag{$\mathbf{P}_k$}$$

where H_k is the Hessian of a Lagrangian for **P** [4]. When the Hessian H_k is used in the subproblems, the method corresponds to a Newton method (4), and when H_k is approximated by a matrix B_k, it corresponds to a quasi-Newton method (5). In either case, the subproblems $\mathbf{P}_k$ may or may not be convex depending on whether H_k, $B_k \succeq 0$. In the context of the broader class of prox-regular h, Lewis and Wright [24] take $B_k = \mu_k I$ at each iteration, thereby guaranteeing existence and uniqueness of the "proximal step" and a global descent algorithm. Instead, our focus is on developing methods possessing fast local rates of convergence by taking advantage of second-order information together with the convex geometry of dom (h) developed by Rockafellar [35].

When h is assumed to be a finite-valued piecewise linear convex function, Womersley [38] established second-order rates of convergence for these algorithms under conditions comparable to those used in NLP, i.e., linear independence of the active constraint gradients, strict complementarity, and strong second-order sufficiency. Notwithstanding this correspondence to NLP, the method of proof differs significantly from the standard methodology for establishing such results in the NLP case developed by Robinson [31, 32]. Notably, in the case of NLP, the function h is piecewise linear but not finite-valued. In subsequent work, Robinson [33] introduced the revolutionary idea of *generalized equations*, whose variational properties can be used to establish local rates of convergence for Newton's method for NLP. By employing the techniques of generalized equations, Cibulka et. al. [8] recently connected classical second-order necessary and sufficient conditions for a local minimizer of **P** with strong metric subregularity (see Definition 9) of the underlying KKT mapping when h is piecewise linear convex but not necessarily finite-valued. However, their analysis relies heavily on the fact that h is piecewise linear. And so, the old question of what conditions imply local quadratic convergence when h is not piecewise linear remains open. However, their technique created the possibility of an extension to the case where h is a member of the PLQ class. This extension is our goal. It is hoped that the methods and techniques developed in this paper provide insight into how to extend these results beyond the PLQ class.

As noted above, we couch the analysis in the context *Newton's method* for generalized equations. The first-order necessary conditions of a local minimum of **P** are encoded through a generalized equation of the form $g(x,y)+G(x,y)\ni 0$, where $g:\mathbb{R}^{n+m}\to\mathbb{R}^{n+m}$ is a $\mathcal{C}^1$-smooth function, $G:\mathbb{R}^{n+m}\rightrightarrows\mathbb{R}^{n+m}$ is a set-valued mapping, (x,y) represents a primal-dual pair, and the function $\nabla g(x,y)$ is a KKT matrix for **P** (see Definition 5). Newton's method (4) for solving this generalized equation corresponds to solving the optimality conditions for $\mathbf{P}_k$. The Newton iterate at (x^k,y^k) is obtained by solving the following linearized generalized equation:

$$\text{Find } (x^{k+1},y^{k+1}) \text{ such that } g(x^k,y^k)+\nabla g(x^k,y^k)\begin{pmatrix}x^{k+1}-x^k\\ y^{k+1}-y^k\end{pmatrix}+G(x^{k+1},y^{k+1})\ni 0. \tag{1}$$

The details of this derivation appear in Section 3.

The goal of this paper is to establish local convergence rates for algorithms based on iteratively solving $\mathbf{P}_k$ in the case where h is a PLQ convex function. We do this by augmenting the strategy of Cibulka et. al. [8] with additional innovations by Lewis [23] and Rockafellar [35]. In particular, we are able to establish conditions under which these algorithms are locally quadratically convergent. The first phase of our analysis involves extensive application of the first- and second-order PLQ calculus [35, 37] to establish conditions under which the underlying generalized equation is strongly metrically subregular. This allows us to establish sufficient conditions for the superlinear convergence of quasi-Newton methods for algorithms whose direction finding subproblems are based on $\mathbf{P}_k$. The second phase of our analysis employs the technique of partly smooth functions in the sense of [20, 23] to establish conditions under which a local approximation to the underlying generalized equation is strongly metrically regular (see Definition 15). This allows us to give conditions for the local quadratic convergence of the Newton method based on $\mathbf{P}_k$.

We also note that recent work by Drusvyatskiy and Lewis [15] considers similar types of results for convex-composite optimization problems of the form $\varphi(x) = h(c(x)) + g(x)$, where h is finite-valued and $L-$Lipschitz, ∇c is β-Lipschitz, and g is closed, proper, convex, and possibly infinite-valued. One of their goals is to understand the convergence of prox-linear type methods through either the subregularity [15, Theorems 5.10 and 5.11] or strong regularity [15, Theorem 6.2] of $\partial\varphi$ at *stable* strong minima or sharp minima of φ [15, Theorems 7.1 and 7.2].

When h is only assumed to be finite-valued convex and g is zero, the first result on the local quadratic convergence for convex-composite problems was that of Burke and Ferris [6]. In that work, the authors established a constraint qualification for the inclusion $c(\overline{x}) \in \arg\min h$ that ensures the local quadratic convergence of constrained Gauss-Newton methods. In [6], the authors assumed $\arg\min h$ was a set of *weak sharp minima* [5]. However, it was observed by Li and Wang [26] that the sharpness hypothesis was not required. Rather, a local quadratic growth condition [26, Theorem 2] was sufficient for the proof techniques in [6] to succeed. The authors continued research [25] in relaxations of the constraint qualification on $c(\overline{x}) \in \arg\min h$ and studied proximal methods [21] for their convergence.

Our focus on the PLQ class is motivated by the great variety of modern problems in data analysis, estimation of dynamical systems, inverse problems, and machine learning that are posed within this class. The key to the success of the convex-composite structure is that it separates the data associated to the problem, the function c, from the model within which we wish to explore the data, the function h. Consequently, the broader the class of functions h available, the greater the variety of ways within which we can explore underlying extremal properties of the input function c, e.g., sparsity, robustness, network structure, dynamics, influence of hyperparameters, etc. Importantly, we have learned that features of the data can be more readily extracted by imposing nonsmoothness in the function h.

The roadmap of the paper is as follows. Section 2 collects tools from convex and variational analysis used throughout the paper. Section 3 formally presents the convex-composite problem class. We take advantage of the structure of the problem class to rewrite the general first-order optimality conditions for proper functions in the presence of various constraint qualifications used in this work. We also present the generalized equation (9) associated with the first-order optimality conditions for $\mathbf{P}$. Section 4 discusses the convex geometry and differential theory of piecewise linear-quadratic functions collected in [37]. The second-order theory of [37] allows us to rewrite the general second-order necessary and sufficient conditions for a local minimum of $\mathbf{P}$. We extract a crucial result from [37] that highlights natural candidates for manifolds of partial smoothness [23] inherent to the function h. Section 5 extends the result [8, Theorem 7.1] relating the strong metric subregularity of (9) to the second-order sufficient conditions of local minima and ends with a convergence study of quasi-Newton methods for $\mathbf{P}$. Section 6 establishes conditions for the partly smooth structure of PLQ convex functions and sets the stage for Section 7, where we analyze the local quadratic convergence of Newton's method as in [13].

2. Notation

These sections summarize the relevant notation and tools of convex and variational analysis used in this work. Unless otherwise stated, we follow the notation in [23, 37, 13].

2.1. Preliminaries

We work in $(\mathbb{R}^n, \langle\cdot,\cdot\rangle)$ with the standard inner product $\langle x, y\rangle = x^\top y = \sum_{i=1}^n x_i y_i$ and $\|x\|^2 = x^\top x$. Throughout, we switch between the notations $\langle x, y\rangle$ and $x^\top y$ for clarity considerations. Let $\mathbb{B} := \left\{x \in \mathbb{R}^n \,\middle|\, \|x\| \le 1\right\}$ be the closed unit ball. For $A \in \mathbb{R}^{m\times n}$, its *range*, *null space*, and *transpose* are $\operatorname{Ran}(A), \operatorname{Null}(A), A^\top$ respectively, and for a finite collection of mappings

$\{A_k\}_{k\in J}$ with index set J, let $\operatorname{diag} A_k$ denote the block diagonal matrix with kth block A_k. Let $e_j \in \mathbb{R}^\ell$ denote the standard unit coordinate vector.

2.2. Convex Analysis A set $C \subset \mathbb{R}^m$ is *locally closed* at a point $\overline{c}$, not necessarily in C, if there exists a closed neighborhood V of $\overline{c}$ such that $C \cap V$ is closed. Any closed set is locally closed at all of its points, and the closure and interior of C is denoted by $\operatorname{cl} C$ and $\operatorname{int} C$, respectively.

For a closed convex set $C \subset \mathbb{R}^m$, let $\operatorname{aff} C$ denote the *affine hull* of C and $\operatorname{par}(C)$ the *subspace parallel* to C. Then, for any $c \in C$, $\operatorname{par}(C) := \operatorname{aff} C - c = \mathbb{R}(C - C)$, where we employ *Minkowski set algebra* for addition of sets: for sets $C_1, C_2 \subset \mathbb{R}^m$ and $t \in \mathbb{R}$, define $C + C' := \left\{ c + c' \mid c \in C,\ c' \in C' \right\}$ and $\Lambda C := \left\{ \lambda c \mid \lambda \in \Lambda,\ c \in C \right\}$. When $C = \{c\}$, we omit the set braces and write $c + C'$. The *relative interior* of C is given by $\operatorname{ri}(C) = \left\{ x \in \operatorname{aff} C \,\middle|\, \exists (\epsilon > 0)\ (x + \epsilon\mathbb{B}) \cap \operatorname{aff} C \subset C \right\}$.

2.3. Variational Analysis The functions in this paper take values in the extended reals $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. For $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the *domain* of f is $\operatorname{dom}(f) := \left\{ x \in \mathbb{R}^n \,\middle|\, f(x) < \infty \right\}$, and the *epigraph* of f is $\operatorname{epi} f := \left\{ (x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \,\middle|\, f(x) \le \alpha \right\}$.
We say f is *closed* if $\operatorname{epi} f$ is a closed subset of $\mathbb{R}^{n+1}$, f is *proper* if $\operatorname{dom}(f) \neq \emptyset$ and $f(x) > -\infty$ for all $x \in \mathbb{R}^n$, and f is *convex* if $\operatorname{epi} f$ is a convex subset of $\mathbb{R}^{n+1}$.

Suppose $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is finite at $\overline{x}$ and $w, v \in \mathbb{R}^n$. The *subderivative* $\mathrm{d}f(\overline{x}) : \mathbb{R}^n \to \overline{\mathbb{R}}$ and *one-sided directional derivative* $f'(\overline{x}; \cdot)$ at $\overline{x}$ for w are

$$\mathrm{d}f(\overline{x})(w) := \liminf_{\substack{t \searrow 0 \\ w' \to w}} \frac{f(\overline{x} + tw) - f(\overline{x})}{t}, \qquad f'(\overline{x}; w) := \lim_{t \searrow 0} \frac{f(\overline{x} + tw) - f(\overline{x})}{t}.$$

At points $w \in \mathbb{R}^n$ such that $f'(\overline{x}; w)$ exists and is finite, the *one-sided second directional derivative* is

$$f''(\overline{x}; w) := \lim_{t \searrow 0} \frac{f(\overline{x} + tw) - f(\overline{x}) - t f'(\overline{x}; w)}{\frac{1}{2} t^2}.$$

For any $w, v \in \mathbb{R}^n$, the *second subderivative at $\overline{x}$ for v and $w \in \mathbb{R}^n$* is

$$\mathrm{d}^2 f(\overline{x}|v)(w) := \liminf_{\substack{t \searrow 0 \\ w' \to w}} \Delta_t^2 f(\overline{x}|v)(w), \text{ where } \Delta_t^2 f(\overline{x}|v)(w) := \frac{f(\overline{x} + tw') - f(\overline{x}) - t\langle v, w' \rangle}{\frac{1}{2} t^2}.$$

The structure of our problem class allows the classical one-sided first and second directional derivatives $f'(\overline{x}; \cdot)$ and $f''(\overline{x}; \cdot)$ to entirely capture the variational properties of their more general counterparts.
Suppose $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is finite at $\overline{x}$. Define the *(Fréchet) regular subdifferential*

$$\widehat{\partial} f(\overline{x}) := \left\{ v \in \mathbb{R}^n \,\middle|\, f(x) \ge f(\overline{x}) + \langle v, x - \overline{x} \rangle + o(\|x - \overline{x}\|) \right\},$$

and the *(limiting or Mordukhovich) subdifferential* by

$$\partial f(\overline{x}) := \left\{ v \in \mathbb{R}^n \,\middle|\, \exists (x^n \underset{f}{\to} \overline{x})\ \exists (v^n \to v)\ \forall (n \in \mathbb{N})\ v^n \in \widehat{\partial} f(x^n) \right\}, \tag{2}$$

where $x^n \underset{f}{\to} \overline{x}$ denotes *f-attentive convergence*, i.e., that $x^n \to \overline{x}$, with $f(x^n) \to f(\overline{x})$. In the case of a closed, proper, convex function f, the set $\partial f(\overline{x})$ is the usual subdifferential of convex analysis. A set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a mapping from $\mathbb{R}^n$ into the power set of $\mathbb{R}^m$, so for each $x \in \mathbb{R}^n$, $S(x) \subset \mathbb{R}^m$. The *graph* and *domain* of S are defined to be

$$\operatorname{gph} S := \left\{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^m \,\middle|\, y \in S(x) \right\} \text{ and } \operatorname{dom}(S) := \left\{ x \in \mathbb{R}^n \,\middle|\, S(x) \neq \emptyset \right\},$$

and S is *graph-convex* whenever $\operatorname{gph} S$ is a convex subset of $\mathbb{R}^n \times \mathbb{R}^m$. For a point $(\overline{x}, \overline{y}) \in \operatorname{gph} S$, and neighborhoods U of $\overline{x}$ and V of $\overline{y}$, a *graphical localization* of S at $\overline{x}$ for $\overline{y}$ is a set-valued mapping $\widetilde{S}$ defined by $\operatorname{gph} \widetilde{S} = \operatorname{gph} S \cap (U \times V)$. A *single-valued localization* of S *at* $\overline{x}$ for $\overline{y}$ is a graphical localization that is also function. If the domain of $\widetilde{S}$ is a neighborhood of $\overline{x}$, $\widetilde{S}$ is called a single-valued localization of S *around* $\overline{x}$ for $\overline{y}$. The mapping S is *outer semicontinuous* at $\overline{x}$ relative to $X \subset \mathbb{R}^n$ if

$$\limsup_{x \xrightarrow[X]{} \overline{x}} S(x) := \left\{ u \;\middle|\; \exists\, (x^n \xrightarrow[X]{} \overline{x})\ \exists\, (u^n \to u)\ \forall\, (n \in \mathbb{N})\ u^n \in S(x^n) \right\} \subset S(\overline{x}),$$

and is *inner semicontinous* relative to $X \subset \mathbb{R}^n$ if

$$S(\overline{x}) \subset \liminf_{x \xrightarrow[X]{} \overline{x}} S(x) := \left\{ u \;\middle|\; \forall\, (x^n \xrightarrow[X]{} \overline{x})\ \exists\, (N \in \mathbb{N},\ u^n \to u)\ \forall\, (n \geq N)\ u^n \in S(x^n) \right\},$$

where $x^n \xrightarrow[X]{} \overline{x} \iff x^n \to \overline{x}$ with $x^n \in X$. Then, (2) is $\partial f(\overline{x}) := \limsup_{x \xrightarrow[f]{} \overline{x}} \widehat{\partial} f(x)$. The last notion employed from variational analysis is that of normal and tangent vectors. Let $C \subset \mathbb{R}^n$, and let $\overline{c} \in C$. Define the *normal cone* to C at $\overline{c}$ as

$$N(\overline{c} \,|\, C) := \limsup_{c \xrightarrow[C]{} \overline{c}} \widehat{N}(c \,|\, C), \text{ where } \widehat{N}(c \,|\, C) := \left\{ v \;\middle|\; \forall\, (c' \in C)\ \left\langle v,\, c' - c \right\rangle \leq o(\left\| c' - c \right\|) \right\}, \tag{3}$$

and the *tangent cone* to C at $\overline{c}$ as $T(\overline{c} \,|\, C) := \limsup_{t \searrow 0} t^{-1}(C - \overline{c})$. A set C is *Clarke regular* at $\overline{c} \in C$ if C is locally closed at $\overline{c}$ and $N(\overline{c} \,|\, C) = \widehat{N}(\overline{c} \,|\, C)$. A nonempty, closed, convex set C is Clarke regular at all $\overline{c} \in C$, with $N(\overline{c} \,|\, C) = \left\{ v \;\middle|\; \langle v, c - \overline{c} \rangle \leq 0 \text{ for all } c \in C \right\}$, and $T(\overline{c} \,|\, C) = \left\{ v \;\middle|\; \langle v, w \rangle \leq 0 \text{ for all } w \in N(\overline{c} \,|\, C) \right\} = \operatorname{cl} \left\{ \mathbb{R}_{++}(C - \overline{c}) \right\}$ [37, Theorem 6.9]. We refer the reader to [37, Chapter 6] for a thorough exposition.

Suppose $g : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^1$-smooth, $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a set-valued mapping with closed graph and $\{\mathbf{B}_k\}_{k \in \mathbb{N}} \subset \mathbb{R}^{m \times n}$. Consider the *generalized equation* $0 \in g(z) + G(z)$. The *Newton method* for $g + G$ is the iteration

$$\text{find } z^{k+1} \text{ such that } 0 \in g(z^k) + \nabla g(z^k)(z^{k+1} - z^k) + G(z^{k+1}), \text{ for } k \in \mathbb{N}, \tag{4}$$

and the *quasi-Newton method* for $g + G$ is the iteration

$$\text{find } z^{k+1} \text{ such that } 0 \in g(z^k) + \mathbf{B}_k(z^{k+1} - z^k) + G(z^{k+1}), \text{ for } k \in \mathbb{N}. \tag{5}$$

3. Convex-composite first- and second-order theory We begin by recalling the basic ingredients of convex-composite optimization and the associated variational structures.

Definition 1 (Convex-composite functions). Let $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ be a closed, proper, convex function and $c : \mathbb{R}^n \to \mathbb{R}^m$ a $\mathcal{C}^2$-smooth function. Define $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ by $f(x) := h(c(x))$. We say the function f is convex-composite.

Definition 2 (Convex-composite Lagrangian). [4] For any $y \in \mathbb{R}^m$, define the function $(yc) : \mathbb{R}^n \to \mathbb{R}$ by $(yc)(x) := \left\langle y,\, c(x) \right\rangle$. The Lagrangian for the convex-composite f is defined by $L(x, y) := (yc)(x) - h^\star(y)$, where $h^\star : \mathbb{R}^m \to \overline{\mathbb{R}}$ denotes the *Fenchel conjugate* of the convex function h defined by $h^\star(y) := \sup_{z \in \mathbb{R}^m} \langle z, y \rangle - h(z)$. The Hessian of L in its first variables is denoted

$$\nabla^2_{xx} L(x, y) = \nabla^2 (yc)(x) = \sum_{i=1}^{m} y_i \nabla^2 c_i(x). \tag{6}$$

Definition 3 (Convex-composite multiplier sets). Suppose f is convex-composite. Define the set of multipliers at $\overline{x} \in \mathrm{dom}\,(f)$ for $v \in \mathbb{R}^n$ as in [37, Theorem 13.14] by

$$Y(\overline{x}, v) := \left\{ y \;\middle|\; \begin{pmatrix} v \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial_x L(\overline{x}, y) \\ \partial_y (-L)(\overline{x}, y) \end{pmatrix} \right\} = \left\{ y \in \partial h(c(\overline{x})) \;\middle|\; \nabla c(\overline{x})^\top y = v \right\}, \tag{7}$$

and define the set of multipliers at $\overline{x}$ for 0 by

$$M(\overline{x}) := Y(\overline{x}, 0) = \mathrm{Null}\left(\nabla c(\overline{x})^\top\right) \cap \partial h(c(\overline{x})). \tag{8}$$

A calculus for convex-composite functions at a point $\overline{x} \in \mathrm{dom}\,(f)$ requires various types of "constraint qualifications." Stronger versions of the *basic constraint qualification* (BCQ) will be employed to ensure uniqueness of the multiplier and underlying strict complementarity properties in later sections.

Definition 4 (Convex-composite constraint qualifications). Suppose f is convex-composite and $\overline{x} \in \mathrm{dom}\,(f)$. We say f satisfies the

- *basic constraint qualification* at $\overline{x}$ if

$$\mathrm{Null}\left(\nabla c(\overline{x})^\top\right) \cap N\left(c(\overline{x}) \,|\, \mathrm{dom}\,(h)\right) = \{0\}, \tag{BCQ}$$

- *transversality condition* at $\overline{x}$ if

$$\mathrm{Null}\left(\nabla c(\overline{x})^\top\right) \cap \mathrm{par}\left(\partial h(c(\overline{x}))\right) = \{0\}, \tag{TC}$$

- *strict criticality condition* at $\overline{x} \in \mathrm{dom}\,(f)$ for $\overline{y}$ if

$$\mathrm{Null}\left(\nabla c(\overline{x})^\top\right) \cap \mathrm{ri}\left(\partial h(c(\overline{x}))\right) = \{\overline{y}\}. \tag{SC}$$

Remark 1. Following [37, Definition 10.23], one says that a convex-composite function f is strongly amenable at $\overline{x} \in \mathrm{dom}\,(f)$ if f satisfies (BCQ) at $\overline{x}$. One says that f is fully amenable at $\overline{x} \in \mathrm{dom}\,(f)$ if f satisfies (BCQ) at $\overline{x}$ and the function h is PLQ convex. Here, we make use of the underlying assumption that c is $\mathcal{C}^2$-smooth.
Notice the basic constraint qualification is a *local property* in the following sense. If f satisfies (BCQ) at $\overline{x}$, then there exists a neighborhood U of $\overline{x}$ such that f satisfies (BCQ) at all $x \in [U \cap c^{-1}(\mathrm{dom}\,(h))]$. Moreover, the basic constraint qualification ensures that the chain rule applies in the subdifferential calculus for convex-composite functions and establishes a foundation for the application of tools from variational analysis.

Theorem 1 **(Convex-composite first-order necessary conditions)**. *Suppose f is convex-composite and $\overline{x} \in \mathrm{dom}\,(f)$ is such that f satisfies* (BCQ) *at $\overline{x}$. Then, $\partial f(\overline{x}) = \nabla c(\overline{x})^\top \partial h(c(\overline{x}))$, and for any $d \in \mathbb{R}^n$,*

$$\mathrm{d}f(\overline{x})(d) = h'(c(\overline{x}); \nabla c(\overline{x})d) = \lim_{t \searrow 0} \frac{\Delta f(\overline{x}; td)}{t},$$

where $\Delta f(\overline{x}; d) := h(c(\overline{x}) + \nabla c(\overline{x})d) - h(c(\overline{x}))$. Suppose, in addition, that $\overline{x}$ is a local solution to **P**. *Then, $M(\overline{x}) := \mathrm{Null}\left(\nabla c(\overline{x})^\top\right) \cap \partial h(c(\overline{x})) \neq \emptyset$, or equivalently, $0 \in \partial f(\overline{x})$, and for any $d \in \mathbb{R}^n$, $h'(c(\overline{x}); \nabla c(\overline{x})d) \geq 0$.*

Proof. This follows from [37, Proposition 8.21, Theorem 10.1, Exercise 10.26(b)]. □
We now establish a relationship between the various notions of a constraint qualification given in Definition 4.

Lemma 1. *Suppose f is convex-composite, $\overline{x} \in \mathrm{dom}\,(f)$, and $\overline{y} \in \mathbb{R}^m$. Then, the following implications hold:*

$$\text{(SC)} \implies \text{(TC)} \implies (M(\overline{x}) = \{\overline{y}\}) \implies \text{(BCQ)}$$

Proof. $[(M(\overline{x})=\{\overline{y}\}) \Longrightarrow (\text{BCQ})]$
Let $M(\overline{x})=\{\overline{y}\}$ and suppose there exists

$$0\neq v\in \text{Null}\left(\nabla c(\overline{x})^\top\right)\cap N\left(c(\overline{x})\,|\,\text{dom}\,(h)\right)\subset \text{Null}\left(\nabla c(\overline{x})^\top\right)\cap \text{par}\left(\partial h(c(\overline{x}))\right).$$

Then, by the subgradient inequality, $v+\overline{y}\in \text{Null}\left(\nabla c(\overline{x})^\top\right)\cap \partial h(c(\overline{x}))=M(\overline{x})$, which is a contradiction.
The rest of the proof follows from the elementary implications

$$(\text{Null}\,(A)\cap \text{ri}\,(C)=\{\overline{y}\}) \Longrightarrow (\text{Null}\,(A)\cap \text{par}\,(C)=\{0\}) \Longrightarrow (\text{Null}\,(A)\cap C=\{\overline{y}\})$$

for closed convex sets C and linear maps A. □
Gauss-Newton methods for iteratively solving **P** are based on finding a search direction that approximates a solution to subproblems of the form

$$\underset{d\in\mathbb{R}^n}{\text{minimize}} \quad h(c(\widehat{x})+\nabla c(\widehat{x})d)+\frac{1}{2}d^\top \widehat{H} d. \tag{$\widehat{\mathbf{P}}$}$$

Local rates of convergence for algorithms of this type, where the function h is assumed to be finite-valued and piecewise linear convex were developed by Womersley [38] based on tools developed for classical nonlinear programming. More recently, Cibulka et. al. [8] successfully applied a modern approach through generalized equations to obtain similar and stronger results again in the piecewise linear convex case. Inspired by these results and the existence of a sophisticated first- and second-order subdifferential calculus for piecewise linear-quadratic convex functions [37], we develop a convergence theory in the piecewise linear-quadratic case from the generalized equations perspective. The basic notational objects for our development are given in the next definition.

Definition 5 (Convex-composite generalized equations). Let f be convex-composite, and define the set-valued mapping $g+G:\mathbb{R}^{n+m}\rightrightarrows\mathbb{R}^{n+m}$ by

$$g(x,y)=\begin{pmatrix}\nabla c(x)^\top y\\ -c(x)\end{pmatrix},\quad G(x,y)=\begin{pmatrix}\{0\}^n\\ \partial h^\star(y)\end{pmatrix}. \tag{9}$$

The associated generalized equation for **P** is $g+G\ni 0$. For a fixed $(\overline{x},\overline{y})\in\mathbb{R}^n\times\mathbb{R}^m$, define the linearization mapping

$$\mathcal{G}:(x,y)\mapsto g(\overline{x},\overline{y})+\nabla g(\overline{x},\overline{y})\begin{pmatrix}x-\overline{x}\\ y-\overline{y}\end{pmatrix}+G(x,y), \tag{10}$$

where $\nabla g(\overline{x},\overline{y})=\begin{pmatrix}\nabla^2_{xx}L(\overline{x},\overline{y}) & \nabla c(\overline{x})^\top\\ -\nabla c(\overline{x}) & 0\end{pmatrix}$.
Observe that for any $\overline{x}\in\text{dom}\,(f)$ where f satisfies (BCQ), $\overline{x}$ satisfies the first-order necessary conditions of Theorem 1 for the problem **P** if and only if there exists $\overline{y}$ such that $(\overline{x},\overline{y})$ solves the generalized equation $g+G\ni 0$. More precisely, we have

$$0\in g(\overline{x},\overline{y})+G(\overline{x},\overline{y})\Leftrightarrow \nabla c(\overline{x})^\top\overline{y}=0 \text{ and } \overline{y}\in\partial h(c(\overline{x}))\Leftrightarrow M(\overline{x})\neq\emptyset. \tag{11}$$

The relationship between the linearization of the generalized equation described in (10) and the subproblems $\widehat{\mathbf{P}}$ is described in the following lemma. The proof follows from Theorem 1.

Lemma 2. *Let f be convex-composite and $(\widehat{x},\widehat{y})\in\mathbb{R}^n\times\mathbb{R}^m$ be such that f satisfies (BCQ) at $\widehat{x}$, and define $\widehat{H}:=\nabla^2_{xx}L(\widehat{x},\widehat{y})$. Then, $(\tilde{d},\tilde{y})\in\mathbb{R}^n\times\mathbb{R}^m$ satisfy the optimality conditions for*

$$\underset{d\in\mathbb{R}^n}{\text{minimize}} \quad h(c(\widehat{x})+\nabla c(\widehat{x})d)+\frac{1}{2}d^\top \widehat{H} d \tag{$\widehat{\mathbf{P}}$}$$

if and only if $(\widehat{x}+\tilde{d},\tilde{y})$ solves the Newton equations for $g+G$: $0\in g(\widehat{x},\widehat{y})+\nabla g(\widehat{x},\widehat{y})\begin{pmatrix}x-\widehat{x}\\ y-\widehat{y}\end{pmatrix}+G(x,y)$.

4. Geometry of PLQ Functions and Their Domains In this section, unless otherwise stated, we let $f := h \circ c$ where h is piecewise linear-quadratic convex and c is $\mathcal{C}^2$-smooth.

Definition 6 (piecewise linear-quadratic). A proper function $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ is called piecewise linear-quadratic (PLQ) if $\operatorname{dom}(h) \neq \emptyset$ and $\operatorname{dom}(h)$ can be represented as the union of $\mathcal{K} \geq 1$ polyhedral sets of the form

$$C_k = \left\{ c \;\middle|\; \left\langle a_{kj}, c \right\rangle \leq \alpha_{kj}, \text{ for all } j \in \{1, \ldots, s_k\} \right\} \tag{12}$$

relative to each of which $h(c)$ is given by an expression of the form $\frac{1}{2}\langle c, Q_k c\rangle + \langle b_k, c\rangle + \beta_k$ for some scalar $\beta_k \in \mathbb{R}$, vector $b_k \in \mathbb{R}^n$, and symmetric matrix Q_k.

Remark 2. The sets C_k do not necessarily form a partition of the set C.
The following lemma is straightforward.

Lemma 3. *Suppose h is piecewise linear-quadratic convex. Then, for any $k \in \mathcal{K}$, the matrices Q_k satisfy $\langle c, Q_k c\rangle \geq 0$ for all $c \in \operatorname{par}(C_k)$.*

Definition 7 (Active indices). For a piecewise linear-quadratic function h and a point $\bar{c} \in \operatorname{dom}(h)$, define the set $\mathcal{K}(\bar{c}) := \left\{ k \in \mathcal{K} \mid \bar{c} \in C_k \right\}$, and write $\overline{k} := |\mathcal{K}(\bar{c})|$, so that $\mathcal{K}(\bar{c}) = \left\{ k_1, k_2, \ldots, k_{\overline{k}} \right\}$.
Our first- and second-order analysis in the PLQ case heavily depends on the following results compiled from [37] into a single proposition for ease of reference.

Proposition 1. *[37, Propositions 10.21, 13.9] If $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ is piecewise linear-quadratic, then $\operatorname{dom}(h)$ is closed and h is continuous relative to $\operatorname{dom}(h)$. Consequently, h is closed. At any point $\bar{c} \in \operatorname{dom}(h)$, $h'(\bar{c};\cdot) = \mathrm{d}h(\bar{c})$, and $h'(\bar{c};\cdot)$ is piecewise linear with $\operatorname{dom}\left(h'(\bar{c};\cdot)\right) = \bigcup_{k \in K(\bar{c})} T\left(\bar{c} \,|\, C_k\right) = T\left(\bar{c} \,|\, \operatorname{dom}(h)\right)$. In particular, for $k \in \mathcal{K}(\bar{c})$ and $w \in T\left(\bar{c} \,|\, C_k\right)$,*

$$h'(\bar{c}; w) = \langle Q_k \bar{c} + b_k, w\rangle . \tag{13}$$

If, in addition, h is convex, then $\operatorname{dom}(h)$ is polyhedral,

$$\emptyset \neq \partial h(\bar{c}) = \bigcap_{k \in \mathcal{K}(\bar{c})} \left\{ y \;\middle|\; y - Q_k \bar{c} - b_k \in N\left(\bar{c} \,|\, C_k\right) \right\}, \tag{14}$$

$h''(\bar{c};\cdot)$ is piecewise linear-quadratic, but not necessarily convex, and for any $w \in \mathbb{R}^m$,

$$0 \leq h''(\bar{c}; w) = \begin{cases} \langle w, Q_k w\rangle & \textit{when } w \in T\left(\bar{c} \,|\, C_k\right), \\ \infty & \textit{when } w \notin T\left(\bar{c} \,|\, \operatorname{dom}(h)\right). \end{cases} \tag{15}$$

For every $y \in \partial h(\bar{c})$, $\mathrm{d}^2 h(\bar{c}|y)$ is piecewise linear-quadratic and convex. Let $K(\bar{c}, y) := \left\{ w \;\middle|\; h''(\bar{c}; w) = \langle y, w\rangle \right\}$. Then, $K(\bar{c}, y)$ is a polyhedral cone, and

$$\mathrm{d}^2 h(\bar{c}|y)(w) = \lim_{\tau \searrow 0} \Delta_\tau^2 h(\bar{c}|y)(w) = \begin{cases} h''(\bar{c}; w) & w \in K(\bar{c}, y), \\ +\infty & \textit{otherwise.} \end{cases} \tag{16}$$

Moreover, there exists a neighborhood V of $\bar{c}$ such that

$$h(c) = h(\bar{c}) + h'(\bar{c}; c - \bar{c}) + \frac{1}{2} h''(\bar{c}; c - \bar{c}) \textit{ for } c \in V \cap \operatorname{dom}(h) . \tag{17}$$

The standard development of first- and second-order optimality conditions requires the notion of directions of non-ascent.

Definition 8 (Directions of non-ascent). Let $f:\mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and $x \in \operatorname{dom}(f)$. The *directions of non-ascent* for f at $\overline{x}$ are denoted by $D(x) := \left\{ d \in \mathbb{R}^n \,\middle|\, \mathrm{d}f(x)(d) \le 0 \right\}$.
By Theorem 1, if f is convex-composite and satisfies (BCQ) at x, then

$$D(x) = \left\{ d \in \mathbb{R}^n \,\middle|\, h'(c(x); \nabla c(x)d) \le 0 \right\}. \tag{18}$$

In the PLQ convex case, (BCQ) ensures that we have the following convenient representation of the set $D(\overline{x})$.

Lemma 4. *Let f be as in* **P**, *and let $\overline{x} \in \mathbb{R}^n$ be such that f satisfies* (BCQ) *at $\overline{x}$. Set $\overline{c} := c(\overline{x})$. Then, $D(\overline{x})$ is convex and the union of finitely many polyhedral closed convex sets with following the representation from Proposition 1:*

$$\begin{aligned} D(\overline{x}) &= \bigcup_{k \in \mathcal{K}(\overline{c})} \left\{ d \,\middle|\, \nabla c(\overline{x})d \in T(\overline{c} \,|\, C_k), \left\langle Q_k \overline{c} + b_k,\, \nabla c(\overline{x})d \right\rangle \le 0 \right\} \\ &= \bigcup_{k \in \mathcal{K}(\overline{c})} \left\{ d \,\middle|\, \begin{array}{l} \left\langle Q_k \overline{c} + b_k,\, \nabla c(\overline{x})d \right\rangle \le 0 \\ \left\langle a_{kj},\, \nabla c(\overline{x})d \right\rangle \le 0, j \in I_k(\overline{c}) \end{array} \right\} \end{aligned} \tag{19}$$

Proof. ($\subset$) Suppose $d \in D(\overline{x})$. By (18), $\nabla c(\overline{x})d \in \operatorname{dom}\left(h'(\overline{c}; \cdot)\right)$. In particular, by Proposition 1, $\nabla c(\overline{x})d \in T(\overline{c} \,|\, C_k)$ for some $k \in \mathcal{K}(\overline{c})$. By (13), we also have $\left\langle Q_k \overline{c} + b_k,\, \nabla c(\overline{x})d \right\rangle = h'(c(\overline{x}); \nabla c(\overline{x})d) \le 0$.
($\supset$) If $d \in \bigcup_{k \in \mathcal{K}(\overline{c})} \left\{ d \,\middle|\, \nabla c(\overline{x})d \in T(c \,|\, C_k), \left\langle Q_k \overline{c} + b_k,\, \nabla c(\overline{x})d \right\rangle \le 0 \right\}$, then for some $k \in \mathcal{K}(\overline{c})$, $\nabla c(\overline{x})d \in T(\overline{c} \,|\, C_k)$. Then, again by Proposition 1, $h'(c(\overline{x}); \nabla c(\overline{x})d) = \left\langle Q_k \overline{c} + b_k,\, \nabla c(\overline{x})d \right\rangle \le 0$, so $d \in D(\overline{x})$. □
We now state the first- and second-order optimality conditions for **P**.

Theorem 2 **(PLQ second-order necessary and sufficient conditions)**. *[35, Theorem 3.4]. Let $h:\mathbb{R}^m \to \overline{\mathbb{R}}$ be piecewise linear-quadratic and convex with $\overline{x} \in \operatorname{dom}(f)$ such that f satisfies* (BCQ) *at $\overline{x}$.*
(a) If f has a local minimum at $\overline{x}$, then $0 \in \nabla c(\overline{x})^\top \partial h(c(\overline{x}))$ and

$$h''(c(\overline{x}); \nabla c(\overline{x})d) + \max\left\{ \left\langle d,\, \nabla^2_{xx} L(\overline{x}, y)d \right\rangle \,\middle|\, y \in M(\overline{x}) \right\} \ge 0$$

for all $d \in D(\overline{x})$.
(b) If $0 \in \nabla c(\overline{x})^\top \partial h(c(\overline{x}))$ and

$$h''(c(\overline{x}); \nabla c(\overline{x})d) + \max\left\{ \left\langle d,\, \nabla^2_{xx} L(\overline{x}, y)d \right\rangle \,\middle|\, y \in M(\overline{x}) \right\} > 0$$

for all $d \in D(\overline{x}) \setminus \{0\}$, then there exists a neighborhood U of $\overline{x}$ and a constant $\gamma > 0$ such that

$$f(x) \ge f(\overline{x}) + \gamma \|x - \overline{x}\|^2 \text{ for all } x \in U \cap \operatorname{dom}(f), \tag{20}$$

i.e., $\overline{x}$ is a strong local minimizer of f.

5. Strong Metric Subregularity of the KKT Mapping In this section we establish conditions under which the set-valued mapping of Definition 5 satisfies strong metric subregularity.

Definition 9 (Strong metric subregularity). A set-valued mapping $S:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *strongly metrically subregular* at $\overline{x}$ for $\overline{y}$ if $(\overline{x}, \overline{y}) \in \operatorname{gph} S$ and there exists $\kappa \ge 0$ and a neighborhood U of $\overline{x}$ such that $\|x - \overline{x}\| \le \kappa \operatorname{dist}\left(\overline{y} \,\middle|\, S(x)\right)$ for all $x \in U$.

Our discussion of strong metric subregularity only requires f to satisfy (BCQ) at $\overline{x} \in \mathrm{dom}\,(f)$.

LEMMA 5. *Consider the KKT mapping $g+G$ and the mapping $\mathcal{G}$ given in Definition 5. Then, strong metric subregularity of $g+G$ at $(\overline{x},\overline{y})$ for 0 is equivalent to the property that $(\overline{x},\overline{y})$ is an isolated point of $\mathcal{G}^{-1}(0)$.*

Proof. By [13, Corollary 3I.10], strong metric subregularity of $g+G$ at $(\overline{x},\overline{y})$ for 0 is equivalent to strong metric subregularity of the linearization $\mathcal{G}$ (10) at $(\overline{x},\overline{y})$.

By [37, Theorem 11.14, Proposition 12.30] the mapping $G(x,y)$ is polyhedral; that is, $\mathrm{gph}\, G$ is the union of finitely many polyhedral sets. Then [13, Corollary 3I.11] establishes the equivalence of strong metric subregularity of $\mathcal{G}$ at $(\overline{x},\overline{y})$ for 0 and $(\overline{x},\overline{y})$ being an isolated point of $\mathcal{G}^{-1}(0)$. □

The main result of this section now follows.

THEOREM 3. *Suppose $h:\mathbb{R}^m \to \overline{\mathbb{R}}$ is piecewise linear-quadratic and convex with $\overline{x} \in \mathrm{dom}\,(f)$ such that f satisfies (BCQ) at $\overline{x}$. Then, the following are equivalent:*
1. *The set $M(\overline{x}) := \mathrm{Null}\left(\nabla c(\overline{x})^\top\right) \cap \partial h(c(\overline{x}))$ in (8) is a singleton and the second-order sufficient conditions of Theorem 2 are satisfied at $\overline{x}$;*
2. *The mapping $g+G$ is strongly metrically subregular at $(\overline{x},\overline{y})$ for 0 and $\overline{x}$ is a strong local minimizer of f.*

Proof. ($\Rightarrow$) By Lemma 5 we argue strong metric subregularity of $g+G$ at $(\overline{x},\overline{y})$ for 0 by showing that there is a neighborhood of $(\overline{x},\overline{y})$ on which $(\overline{x},\overline{y})$ is the unique solution to the generalized equation $\mathcal{G} \ni 0$ (10). After the change of variables $d := x-\overline{x}$, we show that there is a neighborhood U of $(0,\overline{y})$ such that $(d,y)=(0,\overline{y})$ is the unique solution to the generalized equation

$$Hd + \nabla c(\overline{x})^\top y = 0 \tag{21}$$
$$c(\overline{x}) + \nabla c(\overline{x}) d \in \partial h^\star(y) \quad (\Leftrightarrow y \in \partial h(c(\overline{x}) + \nabla c(\overline{x})d)), \tag{22}$$

where $H := \nabla^2_{xx} L(\overline{x},\overline{y})$. Suppose there is no such neighborhood. Then, there exists a sequence of vectors $\{(d^i,y^i)\}_{i\in\mathbb{N}}$ converging to $(0,\overline{y})$ with $(d^i,y^i) \neq (0,\overline{y})$ that solve the generalized equation (21), (22). First assume $d^i \neq 0$ for all $i \in \mathbb{N}$. Define for each $i \in \mathbb{N}$, $t_i := \left\|d^i\right\|$, $v^i := d^i/\left\|d^i\right\|$, and assume without loss of generality that $v^i \to \overline{v}$ and that

$$\left\{c(\overline{x}) + \nabla c(\overline{x}) d^i\right\}_{i\in\mathbb{N}} \subset C_{k_0} \text{ for some } k_0 \in K(c(\overline{x}) + \nabla c(\overline{x}) d^i) \subset K(\overline{c}), \tag{23}$$

since $d^i \to 0$. Taking the inner product on both sides of (21) with d^i, we obtain

$$0 = \left\langle d^i,\, Hd^i\right\rangle + \left\langle d^i,\, \nabla c(\overline{x})^\top y^i\right\rangle \text{ for all } i \in \mathbb{N}. \tag{24}$$

The subgradient inequality for h at $c(\overline{x}) + \nabla c(\overline{x}) d^i$ with subgradient y_i gives

$$\Delta f(\overline{x}; d^i) \le \left\langle d^i,\, \nabla c(\overline{x})^\top y_i\right\rangle = -\left\langle d^i,\, Hd^i\right\rangle. \tag{25}$$

Dividing through (25) by $t_i > 0$ and letting $i \to \infty$, (BCQ), Theorem 1, [37, Proposition 8.21] give

$$\mathrm{d}f(\overline{x})(\overline{v}) = h'(c(\overline{x}); \nabla c(\overline{x})\overline{v}) = \liminf_i \frac{\Delta f(\overline{x}; t_i v^i)}{t_i} \le \lim_i -\left\langle v^i,\, Hd^i\right\rangle = 0,$$

so $\overline{v} \in D(\overline{x}) \setminus \{0\}$. By second-order sufficiency, $h''(c(\overline{x}); \nabla c(\overline{x})\overline{v}) + \overline{v}^\top H \overline{v} > 0$. We now show $\nabla c(\overline{x})\overline{v} \in T\left(\overline{c}\,|\,C_{k_0}\right)$. By (23) and the computation $\frac{c(\overline{x}) + \nabla c(\overline{x})d^i - c(\overline{x})}{t_i} = \nabla c(\overline{x}) v^i \to \nabla c(\overline{x})\overline{v} \in T\left(\overline{c}\,|\,C_{k_0}\right)$. Then by (15), $h''(c(\overline{x}); \nabla c(\overline{x})\overline{v}) = \overline{v}^\top \nabla c(\overline{x})^\top Q_{k_0} \nabla c(\overline{x})\overline{v}$, so that

$$\overline{v}^\top H \overline{v} + \overline{v}^\top \nabla c(\overline{x})^\top Q_{k_0} \nabla c(\overline{x})\overline{v} > 0. \tag{26}$$

On the other hand, by (14),

$$y^i \in \partial h(c(\overline{x})+\nabla c(\overline{x})d^i) = \bigcap_{k\in\mathcal{K}(c(\overline{x})+\nabla c(\overline{x})d^i)} \left\{ y \;\middle|\; y - Q_k(c(\overline{x})+\nabla c(\overline{x})d^i) - b_k \in N\left(c(\overline{x})+\nabla c(\overline{x})d^i \,|\, C_k\right) \right\},$$

and so $y^i - Q_{k_0}(c(\overline{x})+\nabla c(\overline{x})d^i) - b_{k_0} \in N\left(c(\overline{x})+\nabla c(\overline{x})d^i \,|\, C_{k_0}\right)$ for all $i\in\mathbb{N}$. Since $c(\overline{x})\in C_{k_0}$, we have

$$\begin{aligned} 0 &\geq \left\langle y^i - [Q_{k_0}(c(\overline{x})+\nabla c(\overline{x})d^i)+b_{k_0}],\, c(\overline{x}) - [c(\overline{x})+\nabla c(\overline{x})d^i]\right\rangle \\ &= \left\langle y^i - Q_{k_0}(c(\overline{x})+\nabla c(\overline{x})d^i) - b_{k_0},\, -\nabla c(\overline{x})d^i\right\rangle \\ &= -\left\langle d^i,\, \nabla c(\overline{x})^\top y^i\right\rangle + \left\langle Q_{k_0}(c(\overline{x})+\nabla c(\overline{x})d^i)+b_{k_0},\, \nabla c(\overline{x})d^i\right\rangle. \end{aligned}$$

Together with (24),

$$\begin{aligned} 0 &\geq \left\langle d^i,\, Hd^i\right\rangle + \left\langle Q_{k_0}(c(\overline{x})+\nabla c(\overline{x})d^i)+b_{k_0},\, \nabla c(\overline{x})d^i\right\rangle \\ &= \left\langle d^i,\, Hd^i\right\rangle + \left\langle \nabla c(\overline{x})d^i,\, Q_{k_0}\nabla c(\overline{x})d^i\right\rangle + \left\langle Q_{k_0}c(\overline{x})+b_{k_0},\, \nabla c(\overline{x})d^i\right\rangle \\ &= \left\langle d^i,\, Hd^i\right\rangle + \left\langle \nabla c(\overline{x})d^i,\, Q_{k_0}\nabla c(\overline{x})d^i\right\rangle + h'(c(\overline{x});\nabla c(\overline{x})d^i) \quad \text{(by (13))} \\ &\geq \left\langle d^i,\, Hd^i\right\rangle + \left\langle \nabla c(\overline{x})d^i,\, Q_{k_0}\nabla c(\overline{x})d^i\right\rangle, \end{aligned}$$

where the final inequality follows from Theorem 1, Theorem 2, and the observation that $\nabla c(\overline{x})d^i \in C_{k_0} - c(\overline{x}) \subset T\left(c(\overline{x})\,|\,C_{k_0}\right)$. Next, divide the inequality $0 \geq \left\langle d^i,\, Hd^i\right\rangle + \left\langle \nabla c(\overline{x})d^i,\, Q_{k_0}\nabla c(\overline{x})d^i\right\rangle$ by t_i^2 and let $i\to\infty$ to yield the contradiction $0 \geq \overline{v}^\top H\overline{v} + \overline{v}^\top \nabla c(\overline{x})^\top Q_k \nabla c(\overline{x})\overline{v} > 0$.
Consequently, $d^i = 0$ for all i sufficiently large, so without loss of generality, we now suppose $d^i = 0$ for all $i\in\mathbb{N}$. Hence by hypothesis, and $y^i \neq \overline{y}$ for all $i\in\mathbb{N}$. But then we contradict uniqueness of $M(\overline{x})$.
($\Leftarrow$) By Lemma 5, $(\overline{x},\overline{y})$ is an isolated point of $\mathcal{G}^{-1}(0)$. That is, there is a neighborhood U of $(\overline{x},\overline{y})$ on which $(\overline{x},\overline{y})$ is the unique solution to the generalized equation

$$\begin{aligned} &H(x-\overline{x}) + \nabla c(\overline{x})^\top y = 0 \\ &c(\overline{x}) + \nabla c(\overline{x})(x-\overline{x}) \in \partial h^\star(y). \end{aligned}$$

For $x=\overline{x}$, this implies there is a neighborhood $U_{\overline{y}}$ about $\overline{y}$ such that

$$U_{\overline{y}} \cap M(\overline{x}) = \{\overline{y}\}. \tag{27}$$

Suppose there is $y\in M(\overline{x})\setminus U_{\overline{y}}$. Then $y_t = (1-t)\overline{y} + ty \in M(\overline{x})$ for $t\in[0,1]$. But for t small, $y_t \in U_{\overline{y}} \cap M(\overline{x})$, which contradicts (27), so $M(\overline{x})$ is the singleton $\{\overline{y}\}$. Therefore, it only remains to show that the second-order sufficient conditions of Theorem 2 are satisfied at $\overline{x}$.
Since $\overline{x}$ is local minimizer of f at which f satisfies (BCQ), Theorem 1 gives $0\in\nabla c(\overline{x})^\top \partial h(c(\overline{x}))$ and $h'(c(\overline{x});\nabla c(\overline{x})d)\geq 0$ for all $d\in\mathbb{R}^n$. Let $\overline{d}\in\mathbb{R}^n\setminus\{0\}$ with $h'(c(\overline{x});\nabla c(\overline{x})d)=0$, or equivalently, $\overline{d}\in D(\overline{x})$. Without loss of generality, suppose $\left\|\overline{d}\right\| = 1$. In particular, by (19), there exists $k_0\in K(\overline{c})$ such that

$$\nabla c(\overline{x})\overline{d} \in T\left(\overline{c}\,|\,C_{k_0}\right) \text{ and } \left\langle Q_{k_0}\overline{c} + b_{k_0},\, \nabla c(\overline{x})\overline{d}\right\rangle = h'(c(\overline{x});\nabla c(\overline{x})\overline{d}) = 0 \tag{28}$$

Since h is PLQ convex, the second-order necessary conditions of Theorem 2 imply $h''(c(\overline{x});\nabla c(\overline{x})\overline{d}) + \overline{d}^\top H\overline{d} \geq 0$.
We show this inequality is strict to complete the proof. Suppose to the contrary that

$$h''(c(\overline{x});\nabla c(\overline{x})\overline{d}) + \overline{d}^\top H\overline{d} = 0. \tag{29}$$

Then, $\overline{d}\neq 0$ solves the program

$$\begin{array}{ll}\underset{d}{\text{minimize}} & h'(c(\overline{x});\nabla c(\overline{x})d)+\frac{1}{2}h''(c(\overline{x});\nabla c(\overline{x})d)+\frac{1}{2}d^\top Hd\\ \text{subject to} & d\in D(\overline{x}).\end{array}$$

By (17) and continuity of $d\mapsto c(\overline{x})+\nabla c(\overline{x})d$, there exists $\epsilon>0$ so that

$$\Delta f(\overline{x};d)=h'(c(\overline{x});\nabla c(\overline{x})d)+\frac{1}{2}h''(c(\overline{x});\nabla c(\overline{x})d)\text{ for }d\in\epsilon\mathbb{B}\cap\left\{d\;\middle|\;c(\overline{x})+\nabla c(\overline{x})d\in\operatorname{dom}(h)\right\}.$$

By (28) and polyhedrality, $c(\overline{x})+t\nabla c(\overline{x})\overline{d}\in\operatorname{dom}(h)$ for sufficiently small $t>0$. It follows, after shrinking $\epsilon>0$ if necessary, that

$$\Delta f(\overline{x};t\overline{d})+\frac{t^2}{2}\overline{d}^\top H\overline{d}=0\text{ for all }0\le t<\epsilon. \tag{30}$$

Since $0\in\partial f(\overline{x})$ and f satisfies (BCQ) at $\overline{x}$, [37, Equation 13(19)] with $v=0$, $y=\overline{y}$, and $w\in\mathbb{R}^n$ gives $\mathrm{d}^2f(\overline{x}|0)(w)=\mathrm{d}^2\overline{f}(\overline{x}|0)(w)+w^\top Hw$, where $\overline{f}(x):=h(c(\overline{x})+\nabla c(\overline{x})[x-\overline{x}])$ is also piecewise linear-quadratic by the discussion following [37, Equation 13(19)]. Since $\overline{x}$ is a strong local minimizer,

$$\mathrm{d}^2f(\overline{x}|0)(w)=\liminf_{\substack{\tau\searrow 0\\ w'\to w}}\frac{f(\overline{x}+tw')-f(\overline{x})}{\frac{1}{2}\tau^2}\ge\liminf_{\substack{\tau\searrow 0\\ w'\to w}}\gamma\left\|w'\right\|^2=\gamma\|w\|^2.$$

Then, we have $\mathrm{d}^2f(\overline{x}|0)(w)=\mathrm{d}^2\overline{f}(\overline{x}|0)(w)+w^\top Hw\ge\gamma\|w\|^2$. By (16) the lim inf defining $\mathrm{d}^2\overline{f}(\overline{x}|0)(w)$ is also expressed as a limit only in τ (because $\overline{f}$ is piecewise linear-quadratic), so

$$\mathrm{d}^2\overline{f}(\overline{x}|0)(w)=\lim_{\tau\searrow 0}\frac{\overline{f}(\overline{x}+\tau w)-\overline{f}(\overline{x})}{\frac{1}{2}\tau^2}=\lim_{\tau\searrow 0}\frac{\Delta f(\overline{x};\tau w)}{\frac{1}{2}\tau^2}.$$

Putting the last two observations together, $\mathrm{d}^2f(\overline{x}|0)(w)=\lim_{\tau\searrow 0}\frac{\Delta f(\overline{x};\tau w)}{\frac{1}{2}\tau^2}+w^\top Hw\ge\gamma\|w\|^2$. But, for $0<\tau<\epsilon$ and $w=\overline{d}$, (30) gives the contradiction $0=\lim_{\tau\searrow 0}\left\{\frac{\Delta f(\overline{x};\tau\overline{d})+\frac{\tau^2}{2}\overline{d}^\top H\overline{d}}{\frac{1}{2}\tau^2}\right\}=\mathrm{d}^2f(\overline{x}|0)(\overline{d})\ge\gamma\left\|\overline{d}\right\|^2=\gamma>0$. □

As an application of Theorem 3, consider the quasi-Newton method (5) initialized at (x^0,y^0) with $\nabla g(x^k,y^k)$ replaced by

$$\mathbf{B}_k=\begin{pmatrix}B_k & \nabla c(x^k)^\top\\ -\nabla c(x^k) & 0\end{pmatrix}. \tag{31}$$

This choice allows us to relate the quasi-Newton method to the optimality conditions for the subproblems $\widehat{\mathbf{P}}$ through the relation of B_k to $\widehat{H}$, as described in following corollary to [13, Dennis-Moré Theorem for Generalized Equations].

Proposition 2. *Let f be as in* **P**. *Suppose $M(\overline{x})=\{\overline{y}\}$ and the second-order sufficient conditions of Theorem 2 are satisfied at $\overline{x}$. Then, $(\overline{x},\overline{y})$ solves $0\in g(\overline{x},\overline{y})+G(\overline{x},\overline{y})$. Moreover, there exists a neighborhood U of $(\overline{x},\overline{y})$ such that if $(x^0,y^0)\in U$, the sequence $\left\{(x^k,y^k)\right\}_{k\in\mathbb{N}}$ generated from the optimality conditions for*

$$\underset{d\in\mathbb{R}^n}{\text{minimize}}\quad h(c(x^k)+\nabla c(x^k)d)+\frac{1}{2}d^\top B_kd \tag{$\mathbf{Q}_k$}$$

remains in U with $(x^k,y^k)\neq(\overline{x},\overline{y})$ for all $k\in\mathbb{N}$, and

$$(x^k,y^k)\to(\overline{x},\overline{y})\text{ and }(B_k-\nabla^2_{xx}L(x^k,y^k))[x^{k+1}-x^k]=o(||(x^{k+1}-x^k,y^{k+1}-y^k)||),$$

then $(x^k,y^k)\to(\overline{x},\overline{y})$ superlinearly.

Remark 3. Consequently, the sufficient conditions for superlinear convergence of quasi-Newton methods require us to choose B_k as an approximation to the Hessian of the Lagrangian $\nabla^2_{xx} L(x^k, y^k) = \nabla^2(y^k c)(x^k)$ in the update direction $x^{k+1} - x^k$ at every iteration.

We are interested in establishing the local convergence of the Newton iterates associated with the generalized equation (1). This is not given by Proposition 2 and requires an in depth analysis of the local active set identification properties for the PLQ function h. For this purpose, we introduce the notion of partial smoothness.

6. Partial Smoothness

Lewis [23] introduced partial smoothness as a way to generalize classical notions of nondegeneracy, strict complementarity, and active constraint identification by illuminating the appropriate underlying manifold geometry of optimization problems. This allows for a more thorough understanding of the convergence behavior of algorithms applied to nonsmooth optimization problems, where solutions lie on well-defined submanifolds of the parameter space on which the function behaves smoothly and off of which it behaves nonsmoothly. Partial smoothness in the context of **P** allows us in Section 7 to establish metric regularity properties of the solution mapping.

Definition 10. Define a set $\mathcal{M} \subset \mathbb{R}^m$ to be a *manifold* of codimension ℓ around $\bar{c} \in \mathbb{R}^m$ if $\bar{c} \in \mathcal{M}$, and there exists an open set $V \subset \mathbb{R}^m$ containing $\bar{c}$ and a $\mathcal{C}^2$-smooth function $F : V \to \mathbb{R}^\ell$ with surjective derivative throughout V such that $\mathcal{M} \cap V = \{c \in V : F(c) = 0\}$. In which case (see [23]), the *tangent space* to $\mathcal{M}$ at $\bar{c}$ is $T(\bar{c} \,|\, \mathcal{M}) = \mathrm{Null}\left(\nabla F(\bar{c})\right)$, the *normal space* to $\mathcal{M}$ at $\bar{c}$ is $N(\bar{c} \,|\, \mathcal{M}) = \mathrm{Ran}\left(\nabla F(\bar{c})^\top\right)$, both independent of the choice of F. Moreover, the set $\mathcal{M}$ is Clarke regular at $\bar{c}$, and $N(\bar{c} \,|\, \mathcal{M})$ equals the normal cone defined in (3).

Definition 11 (Partial smoothness for closed, convex functions). Suppose $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a closed, proper, convex function and that $\bar{c} \in \mathcal{M} \subset \mathbb{R}^m$. The function h is *partly smooth* at $\bar{c}$ relative to $\mathcal{M}$ if $\mathcal{M}$ is a manifold around $\bar{c}$ and the following four properties hold:

(a) (restricted smoothness) the restriction $h|_{\mathcal{M}}$ is smooth around $\bar{c}$, in that there exists a neighborhood V of $\bar{c}$ and a $\mathcal{C}^2$-smooth function g defined on V such that $h = g$ on $V \cap \mathcal{M}$;
(b) (existence of subgradients) at every point $c \in \mathcal{M}$ close to $\bar{c}$, $\partial h(c) \neq \emptyset$;
(c) (normals and subgradients parallel) $\mathrm{par}\left(\partial h(\bar{c})\right) = N(\bar{c} \,|\, \mathcal{M})$;
(d) (subgradient inner semicontinuity) the subdifferential map ∂h is inner semicontinuous at $\bar{c}$ relative to $\mathcal{M}$.

We say that h is *partly smooth relative to* $\mathcal{M}$ if $\mathcal{M}$ is a manifold and h is partly smooth at each point in $\mathcal{M}$ relative to $\mathcal{M}$.

Remark 4. By [23, Proposition 2.4], requiring (a) - (d) in the definition is equivalent to requiring (a), (b), (d), and *normal sharpness*:

$$h'(\bar{c}; -w) > -h'(\bar{c}; w), \quad \forall\, w \in N(\bar{c} \,|\, \mathcal{M}) \setminus \{0\}, \tag{32}$$

and is also equivalent to requiring (a), (b), (d), and *lineality and tangent equality*:

$$\left\{ w \in \mathbb{R}^m \;\middle|\; -h'(\bar{c}; w) = h'(\bar{c}; -w) \right\} =: \mathrm{lin}\; h'(\bar{c}; \cdot) = T(\bar{c} \,|\, \mathcal{M}). \tag{33}$$

In the context of the PLQ functions given in Definition 6, a natural choice for the active manifold at a point $\bar{c} \in \mathrm{dom}\,(h)$ for **P** is the set given by

$$\mathcal{M}_{\bar{c}} := \mathrm{ri}\left(\bigcap_{k \in \mathcal{K}(\bar{c})} C_k \right), \tag{34}$$

where $\mathcal{K}(\bar{c})$ are the active indices at $\bar{c}$ (see Definition 7). The analysis of the manifold $\mathcal{M}_{\bar{c}}$ requires a more thorough understanding of the structure of $\mathrm{dom}\,(h)$, which we obtain from [37, Lemma

2.50]. It implies that the domain of h has a finite stratification [14, Definition 3.1] for which h is a stratifiable function [14, Definition 3.2]. This stratification is central to our discussion of partial smoothness and is referred to as the Rockafellar-Wets PLQ Representation.

THEOREM 4 **(Rockafellar-Wets PLQ Representation)**. *Suppose h is piecewise linear-quadratic convex and* $\operatorname{int}\operatorname{dom}(h)\neq\emptyset$*. Then, without loss of generality, we may assume the polyhedral sets $\{C_k\}_{k=1}^{\mathcal{K}}$ defining h are given in terms of a common set of $s>0$ hyperplanes $\mathcal{H}:=\left\{(a_j,\alpha_j)\right\}_{j=1}^{s}\subset(\mathbb{R}^m\setminus\{0\})\times\mathbb{R}$, so that for all $k\in\{1,\ldots,\mathcal{K}\}$,*

$$C_k=\left\{c\;\middle|\;\left\langle\omega_{kj}a_j,\,c\right\rangle\le\omega_{kj}\alpha_j,\ \textit{for all } j\in\{1,\ldots,s\}\right\},$$

with $\omega_{kj}\in\{\pm1\}$,

$$I_k(c)=\left\{j\;\middle|\;\left\langle\omega_{kj}a_j,\,c\right\rangle=\omega_{kj}\alpha_j\right\}=\left\{j\;\middle|\;\left\langle a_j,\,c\right\rangle=\alpha_j\right\}\subset\{1,\ldots,s\},\tag{35}$$

and
(a) $\emptyset\neq\operatorname{int}C_k=\left\{c\;\middle|\;\left\langle\omega_{kj}a_j,\,c\right\rangle<\omega_{kj}\alpha_j,\ \textit{for all } j\in\{1,\ldots,s_k\}\right\}$*, for all $k\in\{1,\ldots,\mathcal{K}\}$,*
(b) $\operatorname{int}C_{k_1}\cap\operatorname{int}C_{k_2}=\emptyset$ *when $k_1\neq k_2$.*
Condition (b) implies that if $c\in C_{k_1}\cap C_{k_2}$, then $c\in\operatorname{bdry}C_{k_1}\cap\operatorname{bdry}C_{k_2}$ when $k_1\neq k_2$.

Proof. Following the notation in the proof of [37, Lemma 2.50], for every polyhedron D_j and every $i\in\{1,\ldots,s\}$, either $l_i(x)\le0$ for all $x\in D_j$ or $l_i(x)\ge0$ for all $x\in D_j$. Therefore each affine function is used in the definition of D_j, and D_j is contained entirely within one of the sets C_k, relative to which h takes the form $\frac{1}{2}\langle c,\,Q_kc\rangle+\langle b_k,\,c\rangle+\beta_k$. □

REMARK 5. If all of the polyhedral sets C_k have the same affine hull, or equivalently, if all of their relative interiors are defined with respect to a fixed affine set, then one can translate this fixed affine set to the origin and work entirely within the resulting subspace. In this case, one can replace the interior requirement of Theorem 4 with a relative interior requirement.
The basic assumptions employed for the remainder of this section are listed below.

ASSUMPTION 1.
(a) The function h is PLQ convex with $\operatorname{dom}(h)$ given by the Rockafellar-Wets PLQ representation described in Theorem 4,
(b) $\bar{c}\in\operatorname{dom}(h)$ satisfies $\bar{k}:=|\mathcal{K}(\bar{c})|\ge2$,

REMARK 6. Whenever $\mathcal{K}(\bar{c})=\{k_0\}$, h is continuously differentiable on $\operatorname{int}C_{k_0}$. Therefore, we assume that $\bar{k}\ge2$ and delay the discussion of $\bar{k}=1$ to Section 7.2
The following lemma further supports the choice for the manifold $\mathcal{M}_{\bar{c}}$ as the active manifold.

LEMMA 6. *Let $\mathcal{M}_{\bar{c}}$ be as in (34) and let Assumption 1 hold. Then, for any $c\in\mathcal{M}_{\bar{c}}$, $\mathcal{K}(c)=\mathcal{K}(\bar{c})$, and so $\mathcal{M}_c=\mathcal{M}_{\bar{c}}$. Moreover, for any $k\in\mathcal{K}(\bar{c})$, the active index sets $I_k(c)$ satisfy $I_k(c)=I_k(\bar{c})$*

Proof. Suppose $\mathcal{K}(c)\neq\mathcal{K}(\bar{c})$. Since the definition of $\mathcal{M}_{\bar{c}}$ implies $\mathcal{K}(\bar{c})\subset\mathcal{K}(c)$, there exists $j\in\mathcal{K}(c')\setminus\mathcal{K}(\bar{c})$. By (b) in Theorem 4, we necessarily have $c\in\operatorname{bdry}C_j$.
We first argue the existence of $\epsilon>0$ such that that $(\bar{c}+\epsilon\mathbb{B})\cap C_k=\emptyset$ for all $k\notin\mathcal{K}(\bar{c})$. If no such ϵ exists, since there are only finitely many $k\in K\setminus\mathcal{K}(\bar{c})$, there would exist an index $k_0\notin\mathcal{K}(\bar{c})$ and an infinite sequence $c^n\to\bar{c}$ with $\{c^n\}\subset C_{k_0}$. By closedness of the set C_{k_0}, $\bar{c}\in C_{k_0}$, which is a contradiction.
Since $c,\bar{c}\in\mathcal{M}_{\bar{c}}$, by [36, Theorem 6.4] there exists a $\mu>1$ such that $\tilde{c}:=(1-\mu)\bar{c}+\mu c\in\bigcap_{k\in\mathcal{K}(\bar{c})}C_k$. Since $c\in\operatorname{bdry}C_j$, there exists a $z\in\operatorname{int}C_j$ sufficiently close to c so that the ray $\mathcal{R}:=\left\{\tilde{c}+\lambda(z-\tilde{c})\,|\,0\le\lambda\right\}$ meets $\bar{c}+\epsilon\mathbb{B}$. We consider two cases. To set the stage, for any two points

$x,y\in\mathbb{R}^m$, denote the line segment connecting them by $[x,y]=\{(1-\lambda)x+\lambda y\mid 0\le\lambda\le 1\}$.
Case 1. There is a point $x\in\mathcal{R}\cap(\bar c+\epsilon\mathbb{B})\cap C$. Then $z\in[\tilde c,x]\subset C_k$ for some $k\in\mathcal{K}(\bar c)$. But then $z\in(\operatorname{int}C_j)\cap C_k$, a contradiction.
Case 2. We have $\mathcal{R}\cap(\bar c+\epsilon\mathbb{B})\cap C=\emptyset$. Then there is a point $x\in(\bar c+\epsilon\mathbb{B})\setminus C$ such that $z\in[\tilde c,x]$. Since $x\notin C$, there is a first point, which we denote by $\widehat z$, in C_j on this line segment as one moves from x to $\tilde c$. Then the line segment $[\widehat z,\bar c]\subset C$. The point $\widehat z$ is not on the line segment $[\tilde c,\bar c]$ since then both c' and z would be on the line segment $[\tilde c,\bar c]$ and so $\operatorname{int}C_j\cap\operatorname{bdry}C_k\ne\emptyset$ for some $k\in\mathcal{K}(\bar c)$, a contradiction. Consequently, the points $\tilde c,\bar c$ and $\widehat z$ are not all collinear and hence form a triangle inside of C. Let $\tilde z$ be on the boundary of $\bar c+\epsilon\mathbb{B}$ and on the line segment $[\widehat z,\bar c]$. Then the line segment $[\tilde z,\tilde c]$ passes through $\operatorname{int}C_j$. This is again a contradiction.
Therefore, no such c exists, and $\mathcal{K}(c)=\mathcal{K}(\bar c)$ for all $c\in\mathcal{M}_{\bar c}$.
For the second claim, suppose there exists $k\in\mathcal{K}(\bar c),\ c\in\mathcal{M}_{\bar c}$ and $j\in\{1,\dots,s\}$ with

$$\langle c,\,\omega_{kj}a_j\rangle<\omega_{kj}\alpha_j \text{ and } \langle \bar c,\,\omega_{kj}a_j\rangle=\omega_{kj}\alpha_j. \tag{36}$$

Again by [36, Theorem 6.4], we may choose $\mu>1$ so that $\mu\bar c+(1-\mu)c\in\mathcal{M}_{\bar c}$. In particular, $\mu\bar c+(1-\mu)c\in C_k$. But writing $\mu=1+\epsilon$ with $\epsilon>0$ gives the contradiction

$$\begin{aligned}\omega_{kj}\alpha_j&\ge\langle\mu\bar c+(1-\mu)c,\,\omega_{kj}a_j\rangle\\&=(1+\epsilon)\langle\bar c,\,\omega_{kj}a_j\rangle-\epsilon\langle c,\,\omega_{kj}a_j\rangle>\omega_{kj}\alpha_j \text{ by (36)}.\end{aligned}$$

Therefore $I_k(\bar c)\subset I_k(c)$. Reversing the roles of c and $\bar c$ in (36) gives the other inclusion. □
The previous lemma tells us distinct points $c,c'\in\mathcal{M}_{\bar c}$ have the same active indices $\mathcal{K}(c)$ and $\mathcal{K}(c')$. Moreover, for any active polyhedron C_k, the active hyperplanes for that polyhedron, $I_k(c)$ and $I_k(c')$, at c and c' are the same. This observation offers a global description of $\mathcal{M}_{\bar c}$ in terms of the active hyperplanes at $\bar c$ alone.

Lemma 7. *Let $\mathcal{M}_{\bar c}$ be as in (34), and let Assumption 1 hold. Then,*

$$\mathcal{M}_{\bar c}=\left\{c\;\middle|\;\begin{array}{c}\langle c,\,a_j\rangle=\alpha_j \text{ for all } k\in\mathcal{K}(\bar c),j\in I_k(\bar c)\\ \langle c,\,\omega_{kj}a_j\rangle<\omega_{kj}\alpha_j \text{ for all } k\in\mathcal{K}(\bar c),j\notin I_k(\bar c)\end{array}\right\}.$$

In particular, $I_{k_1}(c)=I_{k_2}(c)$ for all $c\in\mathcal{M}_{\bar c}$ and $k_1,k_2\in\mathcal{K}(\bar c)$. Moreover, for any $k\in\mathcal{K}(\bar c)$ and $c\in\mathcal{M}_{\bar c}$, $T(c\mid\mathcal{M}_{\bar c})=\operatorname{Null}(A_k(\bar c)^\top)$, and $N(c\mid\mathcal{M}_{\bar c})=\operatorname{Ran}(A_k(\bar c))$, where $A_k(\bar c)$ is the matrix whose columns are the gradients of the active constraints at $\bar c\in C_{\bar k}$ in some ordering.

Remark 7. By Lemma 6 and Lemma 7, for all $c\in\mathcal{M}_{\bar c},\ k\in\mathcal{K}(\bar c)$, and $j\in\mathcal{K}(c)$, $\operatorname{Ran}(A_k(\bar c))=\operatorname{Ran}(A_j(c))$. This observation becomes important in a structural definition to follow.

Proof. Define

$$\mathcal{C}_1:=\bigcap_{k\in\mathcal{K}(\bar c)}C_k,\quad \mathcal{C}_2:=\left\{c\;\middle|\;\begin{array}{c}\langle c,\,\omega_{kj}a_j\rangle=\omega_{kj}\alpha_j \text{ for all } k\in\mathcal{K}(\bar c),j\in I_k(\bar c)\\ \langle c,\,\omega_{kj}a_j\rangle\le\omega_{kj}\alpha_j \text{ for all } k\in\mathcal{K}(\bar c),j\notin I_k(\bar c)\end{array}\right\}.$$

We aim to show $\operatorname{ri}(\mathcal{C}_1)\supset\operatorname{ri}(\mathcal{C}_2)$. For $k\in\mathcal{K}(\bar c)$ and $j\in I_k(\bar c)$ define $\mathcal{C}_{k,j}:=\left\{c\;\middle|\;\langle c,\,\omega_{kj}a_j\rangle=\omega_{kj}\alpha_j\right\}$, and for $k\in\mathcal{K}(\bar c)$ and $j\notin I_k(\bar c)$, let $\mathcal{D}_{k,j}:=\left\{c\;\middle|\;\langle c,\,\omega_{kj}a_j\rangle\le\omega_{kj}\alpha_j\right\}$. Then by defintion of $I_k(\bar c)$,

$$\bar c\in\bigcap_{\substack{k\in\mathcal{K}(\bar c)\\ j\in I_k(\bar c)}}\operatorname{ri}(\mathcal{C}_{kj})\cap\bigcap_{\substack{k\in\mathcal{K}(\bar c)\\ j\notin I_k(\bar c)}}\operatorname{ri}(\mathcal{D}_{kj}),$$

so [36, Theorem 6.5] gives

$$\operatorname{ri}(\mathcal{C}_2) = \left\{ c \;\middle|\; \begin{array}{l} \langle c, \omega_{kj} a_j \rangle = \omega_{kj}\alpha_j \text{ for all } k \in \mathcal{K}(\bar{c}), j \in I_k(\bar{c}) \\ \langle c, \omega_{kj} a_j \rangle < \omega_{kj}\alpha_j \text{ for all } k \in \mathcal{K}(\bar{c}), j \notin I_k(\bar{c}) \end{array} \right\}.$$

Moreover, $\mathcal{C}_1 \supset \mathcal{C}_2$ with $\mathcal{C}_2$ not entirely contained within the relative boundary of $\mathcal{C}_1$ because $\bar{c} \in \mathcal{C}_2 \cap \mathcal{M}_{\bar{c}}$. By [36, Corollary 6.5.2], $\mathcal{M}_{\bar{c}} := \operatorname{ri}(\mathcal{C}_1) \supset \operatorname{ri}(\mathcal{C}_2)$. Lemma 6 shows $\mathcal{M}_{\bar{c}} := \operatorname{ri}(\mathcal{C}_1) \subset \operatorname{ri}(\mathcal{C}_2)$ because $I_k(c) = I_k(\bar{c})$ throughout $\mathcal{M}_{\bar{c}}$.
For the second claim, the structure of $\mathcal{M}_{\bar{c}}$ implies that if $\langle c, \omega_{k_1 j} a_j \rangle = \omega_{k_1 j}\alpha_j$ for some $k_1 \in \mathcal{K}(\bar{c})$, then $\langle c, \omega_{k_2 j} a_j \rangle = \omega_{k_2 j}\alpha_j$ for any other $k_2 \in \mathcal{K}(\bar{c})$ as $\omega_{kj} \in \{\pm 1\}$. Hence $I_{k_2}(c) \supset I_{k_1}(c)$, and this argument is symmetric in k_1 and k_2.

The tangent and normal cone formulas hold throughout $\mathcal{M}_{\bar{c}}$ by [37, Theorem 6.46]. □
Based on Lemma 7 and Remark 7, we now establish the notational tools required for our analysis.

DEFINITION 12. Let $\mathcal{M}_{\bar{c}}$ be as in (34), and let Assumption 1 hold. Define $A_{\bar{k}}(c)$ to be the matrix whose columns are the gradients of the active constraints at $c \in C_{\bar{k}}$ in some ordering. By Theorem 4 and Lemma 7, without loss of generality, we can define $A := A_{\bar{k}}(c)$ independent of the choice of $c \in \mathcal{M}_{\bar{c}}$, and for any $j \in \left\{1, \ldots, \bar{k}\right\}$, there exists a diagonal matrix P_j with entries ± 1 on the diagonal such that

$$AP_j = A_{k_j}(c) \text{ independent of } c \in \mathcal{M}_{\bar{c}}. \tag{37}$$

We let ℓ be the common number of columns $\ell := |I_k(\bar{c})| = |I_{k'}(\bar{c})|$ for all $k, k' \in \mathcal{K}(\bar{c})$, so that $A \in \mathbb{R}^{m\times\ell}, P_j \in \mathbb{R}^{\ell\times\ell}, P_{\bar{k}} = I_\ell$, and define the following block matrices $\widehat{\mathcal{Q}} := \operatorname{diag}(Q_k), \widehat{\mathcal{A}} := \operatorname{diag} AP_j$

$$\mathcal{A} := \begin{pmatrix} (1-\bar{k})AP_1 & AP_2 & \cdots & A \\ AP_1 & (1-\bar{k})AP_2 & \cdots & A \\ \vdots & \ddots & \ddots & \vdots \\ AP_1 & AP_2 & \cdots & (1-\bar{k})A \end{pmatrix}, \; \mathcal{Q} := \begin{bmatrix} Q_{k_1} \\ Q_{k_2} \\ \vdots \\ Q_{k_{\bar{k}}} \end{bmatrix}, \; \mathcal{B} := \begin{bmatrix} b_{k_1} \\ b_{k_2} \\ \vdots \\ b_{k_{\bar{k}}} \end{bmatrix}, \; J := \begin{bmatrix} I_m \\ I_m \\ \vdots \\ I_m \end{bmatrix} \tag{38}$$

and averaged quantities

$$\overline{Q} = (1/\bar{k}) J^\top \widehat{\mathcal{Q}} J, \quad \overline{A} = (1/\bar{k}) J^\top \widehat{\mathcal{A}}, \quad \bar{b} = (1/\bar{k}) J^\top \mathcal{B}, \quad \lambda_0(\bar{c}) = \overline{Q}\bar{c} + \bar{b}.$$

In a fashion similar to the *structure functional* approach of [38, 27, 28], we give a formula for the subdifferential in terms of the active manifold structure previously laid out.

LEMMA 8. *Let $\mathcal{M}_{\bar{c}}$ be as in (34), let Assumption 1 hold, and recall the notation of Definition 12. For any $c \in \mathcal{M}_{\bar{c}}$, $\partial h(c)$ can be given by two equivalent formulations:*

$$\partial h(c) = \left\{ y \;\middle|\; \begin{array}{l} \exists\, \mu = (\mu_1^\top, \ldots, \mu_{\bar{k}}^\top)^\top \geq 0 \\ \textit{such that } Jy = \mathcal{Q}c + \mathcal{B} + \widehat{\mathcal{A}}\mu \end{array} \right\} = \lambda_0(c) + \overline{A}\,\mathcal{U}(c), \tag{39}$$

where

$$\mathcal{U}(c) := \left\{ \mu \geq 0 \;\middle|\; \mathcal{A}\mu = \bar{k}\left[\mathcal{Q}c + \mathcal{B} - J(\overline{Q}c + \bar{b})\right] \right\}. \tag{40}$$

Proof. By (14) and Lemma 6, $y \in \partial h(c)$ if and only if $y \in Q_{k_j}c + b_{k_j} + N\left(c \,|\, C_{k_j}\right)$ for all $j \in \left\{1, \ldots, \overline{k}\right\}$. In terms of the active indices at c for the polyhedron C_{k_j}, [37, Theorem 6.46] and (37) imply

$$y = Q_{k_j}c + b_{k_j} + AP_j\mu_j, \text{ where } j \in \left\{1, \ldots, \overline{k}\right\}, \mu_j \geq 0.$$

Hence $y \in \partial h(c)$ if and only if there exists $\mu = (\mu_1^\top, \ldots, \mu_{\overline{k}}^\top)$ such that (y, μ) satisfies the system

$$Jy = \mathcal{Q}c + \mathcal{B} + \widehat{\mathcal{A}}\mu, \quad \mu = (\mu_1^\top, \ldots, \mu_{\overline{k}}^\top)^\top \geq 0.$$

Since $J^\top J = \overline{k}I_m$, multiplying both sides of the first equation in (39) by $(1/\overline{k})J^\top$ gives $y = \overline{Q}c + \overline{b} + \overline{A}\mu$, where μ satisfies

$$\overline{Q}c + \overline{b} + \overline{A}\mu = AP_j\mu_j + Q_{k_j}c + b_{k_j}, \text{for all } j \in \left\{1, \ldots, \overline{k}\right\}, \mu \geq 0.$$

The set of μ that satisfy the display defines membership in $\mathcal{U}(c)$, so $\partial h(c) = \lambda_0(c) + \overline{A}\mathcal{U}(c)$. □

The notion of nondegeneracy that we use imposes linear independence of the columns of A.

Definition 13 (Nondegeneracy). Let $\mathcal{M}_{\overline{c}}$ be as in (34), let Assumption 1 hold, and recall the notation of Definition 12. We say that $\mathcal{M}_{\overline{c}}$ satisfies the *nondegeneracy condition* if $\mathrm{Null}(A) = \{0\}$.

Nondegenercy yields a uniqueness property of the multipliers $\mu \in \mathcal{U}(c)$.

Lemma 9. *Let $\mathcal{M}_{\overline{c}}$ be as in (34), let Assumption 1 hold, and recall the notation of Definition 12. Suppose $\mathcal{M}_{\overline{c}}$ satisfies the nondegeneracy condition of Definition 13, $c \in \mathcal{M}_{\overline{c}}$, and $y \in \partial h(c)$. Then, there is a unique $\mu \in \mathcal{U}(c)$, given by $\mu(c,y)_j = P_j(A^\top A)^{-1}A^\top(y - (Q_{k_j}c + b_{k_j})), j \in \left\{1, \ldots, \overline{k}\right\}$ so that $y = \lambda_0(c) + \overline{A}\mu(c,y)$.*

Proof. For any $j \in \left\{1, \ldots, \overline{k}\right\}$, Lemma 8 implies there exists $\mu_j \geq 0$ such that $y = Q_{k_j}c + b_{k_j} + AP_j\mu_j$. Nondegeneracy implies μ_j is given uniquely by the equation $\mu(c,y)_j = P_j(A^\top A)^{-1}A^\top(y - (Q_{k_j}c + b_{k_j}))$. □

A corresponding notion of strict complementarity is provided by the next lemma.

Lemma 10. *Let $\mathcal{M}_{\overline{c}}$ be as in (34), let Assumption 1 hold, and recall the notation of Definition 12. Suppose $c \in \mathcal{M}_{\overline{c}}$ and $\mathrm{ri}\left(\partial h(c)\right) \neq \emptyset$. Then $y \in \mathrm{ri}\left(\partial h(c)\right)$ if and only if $\mu(c,y)_i > 0$ for all $i \in \left\{1, \ldots, \overline{k}\right\}$.*

Proof. By [36, Theorem 6.4], $y \in \mathrm{ri}\left(\partial h(c)\right)$ if and only if for all $y' \in \partial h(c)$, there exists $t > 1$ so that $ty + (1-t)y' \in \partial h(c)$. Choose a $y' \in \partial h(c)$ with $y' \neq y$.

($\Rightarrow$) If there exists $i_0 \in \left\{1, \ldots, \overline{k}\right\}$ and $j \in \{1, \ldots, \ell\}$, with $(\mu(c,y)_{i_0})_j = 0$, then, by (39),

$$\partial h(c) \ni ty + (1-t)y' = Q_{i_0}c + b_{i_0} + AP_{i_0}[t\mu(c,y)_{i_0} + (1-t)\mu(c,y')_{i_0}].$$

By Lemma 9, $\mu(c, ty + (1-t)y')_{i_0} = t\mu(c,y)_{i_0} + (1-t)\mu(c,y')_{i_0}$. By assumption, the right-hand side has its jth component is negative for all $t > 1$, a contradiction.

($\Leftarrow$) We must show there exists $\epsilon > 0$ such that if $t := 1 + \epsilon$ then $t\mu(c,y)_{i_0} + (1-t)\mu(c,y')_{i_0} > 0$. After rearranging, this is equivalent to finding $\epsilon > 0$ so that $\mu(c,y)_{i_0} + \epsilon[\mu(c,y)_{i_0} - \mu(c,y')_{i_0}] > 0$. If $\mu(c,y)_{i_0} - \mu(c,y')_{i_0} \geq 0$, the claim is immediate. Otherwise, we choose ϵ via

$$0 < \epsilon < \min\left\{ \frac{(\mu(c,y)_{i_0})_j}{(\mu(c,y')_{i_0})_j - (\mu(c,y)_{i_0})_j} \,\Big|\, (\mu(c,y)_{i_0})_j - (\mu(c,y')_{i_0})_j < 0, j \in \{1, \ldots, \ell\} \right\}.$$

Then $y \in \mathrm{ri}\left(\partial h(c)\right)$. □

However, a weaker notion of strict complementarity in conjunction with nondegeneracy suffices to show that $\mathrm{ri}\left(\partial h(c)\right) \neq \emptyset$ throughout $\mathcal{M}_{\overline{c}}$.

Definition 14 (k-strict complementarity). Let $\mathcal{M}_{\overline{c}}$ be as in (34), let Assumption 1 hold, and recall the notation of Definition 12. We say *k-strict complementarity* holds at (c,y) for $\mu = (\mu_1^\top, \dots, \mu_{\overline{k}}^\top)^\top$ if
(a) $c \in \mathcal{M}_{\overline{c}}$, $y \in \partial h(c)$,
(b) There exists $k \in \mathcal{K}(\overline{c})$ with $\mu_k > 0$,
(c) Whenever there exists $j \in \mathcal{K}(c) \setminus \{k\}$ and $i \in \{1, \dots, \ell\}$ with $(\mu_j)_i = 0$, then the scalars $(P_{j'})_{ii} = 1$ for all $j' \in \mathcal{K}(c)$,
(d) (y, μ) satisfies (39).

Remark 8. When k-strict complementarity holds at a pair (c,y) and an index j satisfies (c), the active polyhedra $\{C_k\}_{k \in \mathcal{K}(\overline{c})}$ are all within the same closed half-space of the corresponding hyperplane. Also observe that $y \in \operatorname{ri}\left(\partial h(c)\right)$ implies k-strict complementarity at (c,y).
A requirement of partial smoothness is that the normal space to $\mathcal{M}_{\overline{c}}$ and $\operatorname{par}\left(\partial h(c)\right)$ are equal. The nondegeneracy condition allows us to describe $\operatorname{par}\left(\partial h(c)\right)$ using the vectors in $\mathcal{U}(c)$ rather than the subgradients in $\partial h(c)$.

Lemma 11. *Let $\mathcal{M}_{\overline{c}}$ be as in (34), let Assumption 1 hold, and recall the notation of Definition 12. Suppose $\mathcal{M}_{\overline{c}}$ satisfies the nondegeneracy condition. Then, for any $c \in \mathcal{M}_{\overline{c}}$,*

$$\operatorname{par}\left(\partial h(c)\right) = \operatorname{Ran}(A) \iff \operatorname{par}\left(\mathcal{U}(c)\right) = \operatorname{Null}(\mathcal{A}). \tag{41}$$

Proof. By Lemma 7, $N(c \,|\, \mathcal{M}_{\overline{c}}) = \operatorname{Ran}(A)$, and by Lemma 8, $\partial h(c) = \lambda_0(c) + \overline{A}\mathcal{U}(c)$. The system of linear equations (40) in $\mathcal{U}(c)$ has coefficient matrix $\mathcal{A}$ defined in (38) which is block-circulant and can be block row-reduced to

$$\begin{pmatrix} AP_1 & 0 & 0 & \cdots & -A \\ 0 & AP_2 & 0 & \cdots & -A \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & AP_{\overline{k}-1} & -A \\ 0 & \cdots & \cdots & 0 & 0 \end{pmatrix}. \tag{42}$$

We now compute $\operatorname{Null}(\mathcal{A})$. Suppose $\mu = (\mu_1^\top, \dots, \mu_{\overline{k}}^\top)^\top \in \operatorname{Null}(\mathcal{A})$. Then (42) and nondegeneracy imply that $\mu \in \operatorname{Null}(\mathcal{A})$ if and only if $\mu_j = P_j \mu_{\overline{k}}$ for all $j \in \left\{1, \dots, \overline{k}-1\right\}$, i.e.,

$$\operatorname{Null}(\mathcal{A}) = \left\{ \begin{pmatrix} P_1 \mu_{\overline{k}} \\ \vdots \\ P_{\overline{k}-1}\mu_{\overline{k}} \\ \mu_{\overline{k}} \end{pmatrix} \,\middle|\, \mu_{\overline{k}} \in \mathbb{R}^\ell \right\}, \text{with basis } \left\{ \begin{pmatrix} P_1 e_p \\ \vdots \\ P_{\overline{k}-1} e_p \\ e_p \end{pmatrix} \,\middle|\, p \in \{1, \dots, \ell\} \right\} =: \{\zeta_1, \dots, \zeta_\ell\}. \tag{43}$$

By (40),

$$\operatorname{par}\left(\mathcal{U}(c)\right) := \mathbb{R}(\mathcal{U}(c) - \mathcal{U}(c)) \subset \operatorname{Null}(\mathcal{A}), \tag{44}$$

and since $\overline{A} = \frac{1}{\overline{k}}\left[AP_1 \, \cdots \, AP_{\overline{k}-1} \; A\right]$, (39) implies

$$\operatorname{par}\left(\partial h(c)\right) = \operatorname{par}\left(\overline{A}\mathcal{U}(c)\right) = \overline{A}\operatorname{par}\left(\mathcal{U}(c)\right) \subset \overline{A}\operatorname{Null}(\mathcal{A}) = \left\{ A\mu_k \,\middle|\, \mu_k \in \mathbb{R}^\ell \right\} = \operatorname{Ran}(A),$$

so ($\Leftarrow$) in (41) is clear as "$\subset$" becomes an equation. For ($\Rightarrow$), suppose strict containment: $\operatorname{par}\left(\mathcal{U}(c)\right) \subsetneq \operatorname{Null}(\mathcal{A})$. Then there exists $p \in \{1, \dots, \ell\}$ such that $\zeta_p \notin \operatorname{par}\left(\mathcal{U}(c)\right)$. This implies that the pth column of A is not in $\operatorname{par}\left(\partial h(c)\right)$ which we have assumed equal to $\operatorname{Ran}(A)$. This contradiction establishes (41). □

We now show that nondegeneracy and k-strict complementarity together imply that the normal space and subdifferential are parallel.

Lemma 12. *Let $\mathcal{M}_{\overline{c}}$ be as in (34), let Assumption 1 hold, and recall the notation of Definition 12. Suppose $\mathcal{M}_{\overline{c}}$ satisfies the nondegeneracy condition, and the k-strict complementarity of Definition 14 holds at (c, y) for μ. Then,*

$$\operatorname{par}\big(\partial h(c)\big) = N\left(c \,|\, \mathcal{M}_{\overline{c}}\right), \tag{45}$$

where it is shown in Lemma 7 that $N(c \,|\, \mathcal{M}_{\overline{c}}) = \operatorname{Ran}(A)$. Moreover, (45) holds throughout $\mathcal{M}_{\overline{c}}$, and ∂h is inner semicontinuous relative to $\mathcal{M}_{\overline{c}}$.

Proof. We first show that a sufficient condition to guarantee the right-hand side of (41) is (c, v) satisfying the k-strict complementarity condition of Definition 14 for $\mu \in \mathcal{U}(c)$. To see this note that, by relabeling the active polyhedral sets if necessary, we can assume without loss of generality that the index k in k-strict complementarity is $\overline{k}$. Let $p \in \{1, \ldots, \ell\}$, $t \in \mathbb{R}$, and consider the step given by $\mu + t\zeta_p$, where ζ_p is the pth basis element of $\operatorname{Null}(\mathcal{A})$ given in (43), i.e.,

$$\mu + t\zeta_p := \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_{\overline{k}-1} \\ \mu_{\overline{k}} \end{pmatrix} + t \begin{pmatrix} P_1 e_p \\ \vdots \\ P_{\overline{k}-1} e_p \\ e_p \end{pmatrix}, \tag{46}$$

We consider two cases. If, for all $j \in \left\{1, \ldots, \overline{k}\right\}$, $(\mu_j)_p > 0$, then for sufficiently small t, $\mu + t\zeta_p \geq 0$, and $\mathcal{A}(\mu + t\zeta_p) = \mathcal{A}\mu$. That is, both $\mu \in \mathcal{U}(c)$ and $\mu + t\zeta_p \in \mathcal{U}(c)$, which implies $\zeta_p \in \operatorname{par}\big(\mathcal{U}(c)\big)$. Otherwise, there exists $j \in \left\{1, \ldots, \overline{k}\right\}$ with $(\mu_j)_p = 0$. By part (c) of k-strict complementarity, the scalars $P_{j'} e_p = 1$ for all $j' \in \left\{1, \ldots, \overline{k}\right\}$, so repeating the previous argument with $t > 0$ gives $\zeta_p \in \operatorname{par}\big(\mathcal{U}(c)\big)$. Since $p \in \{1, \ldots, \ell\}$ was arbitrary, k-strict complementarity is a sufficient condition guaranteeing $\operatorname{par}\big(\mathcal{U}(c)\big) = \operatorname{Null}(\mathcal{A})$.
This argument shows, under nondegeneracy, that

$$k\text{-strict complementarity at } (c, y) \text{ for } \mu \Longrightarrow \operatorname{ri}\big(\partial h(c)\big) \neq \emptyset, \tag{47}$$

because, given any $\mu \in \mathcal{U}(c)$, the fact that $\operatorname{par}\big(\mathcal{U}(c)\big) = \operatorname{Null}(\mathcal{A})$ together with (39) implies there exists a strictly positive $\tilde{\mu} \in \mathcal{U}(c)$ and a $\tilde{y} \in \partial h(c)$ given by $\tilde{y} = \lambda_0(c) + \overline{A}\tilde{\mu}$, with $\mu(c, \tilde{y}) = \tilde{\mu}$. By Lemma 10, $\tilde{y} \in \operatorname{ri}\big(\partial h(c)\big)$.
We now argue that if, for some $c \in \mathcal{M}_{\overline{c}}$, $y \in \partial h(c)$, k-strict complementarity holds at (c, y) for μ, then $\operatorname{ri}\big(\partial h(c)\big) \neq \emptyset$ throughout $\mathcal{M}_{\overline{c}}$. This will imply (45) holds throughout $\mathcal{M}_{\overline{c}}$ as well. By (47), suppose $y \in \operatorname{ri}\big(\partial h(c)\big)$ so that $\mu(c, y) > 0$ by Lemma 10.
Choose any other $c' \in \mathcal{M}_{\overline{c}}$. Since $\mathcal{M}_{\overline{c}}$ is relatively open, there exists $c'' \in \mathcal{M}_{\overline{c}}$ and $\lambda \in (0, 1)$ so that $c' = \lambda c + (1 - \lambda)c''$. Let $y'' \in \partial h(c'')$. By Lemma 9, there exists a unique vector $\mu(c'', y'')$ associated with (c'', y''). Since $c,\ c'' \in \mathcal{M}_{\overline{c}}$ and $\mu(c, y) > 0$, $\lambda\mu(c', y') + (1 - \lambda)\mu(c, y) > 0$. It follows from (39) that for all $j \in \left\{1, \ldots, \overline{k}\right\}$ and $\lambda \in (0, 1)$,

$$\lambda y + (1 - \lambda)y'' = Q_{k_j} c' + b_{k_j} + A P_j(\lambda\mu(c, y) + (1 - \lambda)\mu(c'', y'')). \tag{48}$$

Define $y' := \lambda y + (1 - \lambda)y''$. Then (48) implies that the equations (39) defining membership $y' \in \partial h(c')$ are satisfied, with $\mu(c', y') = \lambda\mu(c, y) + (1 - \lambda)\mu(c'', y'') > 0$, so $y' \in \operatorname{ri}\big(\partial h(c')\big)$ by Lemma 10. Since $c' \in \mathcal{M}_{\overline{c}}$ was arbitrary, $\operatorname{ri}\big(\partial h(c)\big) \neq \emptyset$ for all $\mathcal{M}_{\overline{c}}$.
We lastly establish $\partial h(c)$ is inner semicontinuous relative to $\mathcal{M}_{\overline{c}}$. The previous paragraph and (48) showed $\partial h|_{\mathcal{M}_{\overline{c}}}$ is graph-convex. By defining $S(c) = \partial h(c)$ for $c \in \mathcal{M}_{\overline{c}}$ and $S(c) = \emptyset$ otherwise and noting the convex sets $\{c\}$ and $\mathcal{M}_{\overline{c}}$ cannot be separated, [37, Theorem 5.9(b)] gives inner semicontinuity of ∂h at all $c \in \mathcal{M}_{\overline{c}}$ relative to $\mathcal{M}_{\overline{c}}$. □
The main result of this section shows that partial smoothness follows from nondegeneracy and k-strict complementarity.

Theorem 5. *Let $\mathcal{M}_{\overline{c}}$ be as in (34), let Assumption 1 hold, and recall the notation of Definition 12. Suppose $\mathcal{M}_{\overline{c}}$ satisfies the nondegeneracy condition, and $c \in \mathcal{M}_{\overline{c}}$ and $y \in \partial h(c)$ are such that (c, y) satisfies the k-strict complementarity condition of Definition 14. Then h is partly smooth relative to $\mathcal{M}_{\overline{c}}$.*

Proof. By definition of $\mathcal{M}_{\overline{c}}$, for any $k \in \mathcal{K}(\overline{c})$ and any $c \in \mathcal{M}_{\overline{c}}$, $h(c) = \frac{1}{2}\langle c, Q_k c\rangle + \langle b_k, c\rangle + \beta_k$, so $h|_{\mathcal{M}_{\overline{c}}}$ is smooth. By Proposition 1, $\mathrm{dom}\,(\partial h) = \mathrm{dom}\,(h) \supset \mathcal{M}_{\overline{c}}$, so existence of subgradients holds throughout $\mathcal{M}_{\overline{c}}$ as well. The normal cone and subdifferential being parallel along with subdifferential inner semicontinuity relative to $\mathcal{M}_{\overline{c}}$ are the content of Lemma 12. □

Remark 9. Observe that if the hypotheses of Theorem 5 are satisfied, the assumption that f satisfies (TC) at $\overline{x}$ is equivalent to requiring

$$\mathrm{Null}\left(\nabla c(\overline{x})^\top\right) \cap \mathrm{Ran}\,(A) = \{0\}. \tag{49}$$

This condition and the nondegeneracy condition imply the $n \times \ell$ matrix $\nabla c(\overline{x})^\top A$ has full rank equal to $\ell \le n$, i.e., $\mathrm{Null}\left(\nabla c(\overline{x})^\top A\right) = \{0\}$.
We now show the assumptions of Theorem 5 allow us to write the cone of non-ascent directions as a subspace at strictly critical points.

Lemma 13 **(Non-ascent directions)**. *Let $\mathcal{M}_{\overline{c}}$ be as in (34), let Assumption 1 hold, and recall the notation of Definition 12. Suppose f satisfies (BCQ) at $\overline{x}$, $\overline{y} \in M(\overline{x})$, and $\overline{c} := c(\overline{x})$. Then, $D(\overline{x}) \supset \mathrm{Null}\left(A^\top \nabla c(\overline{x})\right)$. If, in addition, f satisfies (SC) at $\overline{x}$ for $\overline{y}$ and $\mathcal{M}_{\overline{c}}$ satisfies the nondegeneracy condition, then $D(\overline{x}) \subset \mathrm{Null}\left(A^\top \nabla c(\overline{x})\right)$.*

Proof. Since f satisfies (BCQ) at $\overline{x}$, Theorem 1 gives $D(\overline{x}) = \left\{ d \in \mathbb{R}^n \;\middle|\; h'(c(\overline{x}); \nabla c(\overline{x})d) \le 0 \right\}$.
($\supset$) Since $\overline{y} \in M(\overline{x})$, by (39), there exists $\overline{\mu} \in \mathcal{U}(\overline{c})$ so that $J\overline{y} = \mathcal{Q}\overline{c} + \mathcal{B} + \widehat{\mathcal{A}}\overline{\mu}$. Then, for any $j \in \left\{1, \ldots, \overline{k}\right\}$,

$$\begin{aligned}
D(\overline{x}) &= \bigcup_{j=1}^{\overline{k}} \left\{ d \;\middle|\; \begin{array}{l} \left\langle Q_{k_j}\overline{c} + b_{k_j},\, \nabla c(\overline{x})d \right\rangle \le 0 \\ P_j A^\top \nabla c(\overline{x}) d \le 0 \end{array} \right\} && \text{by (19), Definition 12} \\
&= \bigcup_{j=1}^{\overline{k}} \left\{ d \;\middle|\; \begin{array}{l} \left\langle \overline{y} - AP_j\overline{\mu}_j,\, \nabla c(\overline{x})d \right\rangle \le 0 \\ P_j A^\top \nabla c(\overline{x}) d \le 0 \end{array} \right\} && \text{since } \overline{y} \in M(\overline{x}) \\
&= \bigcup_{j=1}^{\overline{k}} \left\{ d \;\middle|\; \begin{array}{l} \left\langle \overline{\mu}_j,\, P_j A^\top \nabla c(\overline{x})d \right\rangle \ge 0 \\ P_j A^\top \nabla c(\overline{x}) d \le 0 \end{array} \right\}.
\end{aligned}$$

The inclusion follows.
($\subset$) Let $0 \ne d \in D(\overline{x})$, and suppose to the contrary that $d = d_1 + d_2$, where $d_1 \in \mathrm{Null}\left(A^\top \nabla c(\overline{x})\right)$ and $d_2 = \nabla c(\overline{x})^\top A w$, $w \ne 0$. By Lemma 12, $\mathrm{Ran}\,(A) \subset \mathrm{par}\left(\partial h(\overline{c})\right)$. Since $\overline{y} \in \mathrm{ri}\left(\partial h(\overline{c})\right)$, there exists $\epsilon > 0$ so that $\overline{y} + \epsilon A w \in \partial h(\overline{c})$. Then,

$$\begin{aligned}
0 &\ge h'(c(\overline{x}); \nabla c(\overline{x})d) \\
&= \sup_{y \in \partial h(\overline{c})} \left\langle \nabla c(\overline{x})^\top y,\, d \right\rangle \\
&\ge \left\langle \overline{y} + \epsilon A w,\, \nabla c(\overline{x})(d_1 + \nabla c(\overline{x})^\top A w) \right\rangle
\end{aligned}$$

$$\geq \left\langle \nabla c(\overline{x})^\top \overline{y},\, d \right\rangle + \epsilon \left\| \nabla c(\overline{x})^\top A w \right\|^2$$
$$= \epsilon \left\| \nabla c(\overline{x})^\top A w \right\|^2,$$

so $w = 0$ (see Remark 9). □

By a continuity argument in (x, y), we have the following result which is important for our discussion of the metric regularity of Newton's iteration in the next section. It states that, in the presence of partial smoothness, (TC) and the curvature condition are local properties.

LEMMA 14. *Suppose* (49) *holds and that for all* $j \in \mathcal{K}(\overline{c})$ *and*

$$d^\top \nabla c(\overline{x})^\top Q_j \nabla c(\overline{x}) d + d^\top \nabla^2_{xx} L(\overline{x}, \overline{y}) d > 0, \quad \forall\, d \in \mathrm{Null}\left(A^\top \nabla c(\overline{x})\right) \setminus \{0\}.$$

Then, there exists a neighborhood $\mathcal{N}$ *of* $(\overline{x}, \overline{y})$ *such that if* $(x, y) \in \mathcal{N}$ *then for all* $j \in \mathcal{K}(\overline{c})$,

$$d^\top \nabla c(x)^\top Q_j \nabla c(x) d + d^\top \nabla^2_{xx} L(x, y) d > 0, \quad \forall\, d \in \mathrm{Null}\left(A^\top \nabla c(x)\right) \setminus \{0\}, \tag{50}$$

and $\mathrm{Null}\left(\nabla c(x)^\top\right) \cap \mathrm{Ran}\,(A) = \{0\}$.

The following examples are inspired by the discussion in [23].

EXAMPLE 1. In $\mathbb{R}^2$, let $h_a(c) = \|c\|_1^2$, so h is piecewise linear-quadratic convex. If $\mathcal{M} := \{0\}$, then h_a is not partly smooth relative to $\mathcal{M}$ because $\partial h_a(0) = \{0\}$ while $N(0 \,|\, \mathcal{M}) = \mathbb{R}^n$. On the other hand, if $h_b(c) = \|c\|_1$ with the same domain representation, then $\partial h(0) = \mathbb{B}_\infty$, in which case h_b is partly smooth relative to $\mathcal{M}$.

Suppose we represent the domain of h_a and h_b as the four quadrants in the plane, relative to each of which h_a, h_b are linear-quadratic. This representation meets the criteria of the Rockafellar-Wets PLQ representation of Theorem 4. For both h_a and h_b, the nondegeneracy condition for $\mathcal{M}$ holds since A can be taken to be I_2.

EXAMPLE 2. In $\mathbb{R}^2$, the domain of h_a and h_b in the previous example can be presented in the following way. Take each of the four quadrants in the plane and split them along their respective diagonal. Define h_a as usual on each of the pieces. Then this presentation describes $\mathrm{dom}\,(h_a)$ using 4 hyperplanes and also meets the Rockafellar-Wets PLQ representation theorem. However, the nondegeneracy condition fails for $\mathcal{M}$ in this representation.

On the manifold $\mathcal{M}$ given by an "artificial" diagonal, the matrix A is comprised of a single column, with $N(c \,|\, \mathcal{M}) = \mathrm{Ran}\,(A)$ for any $c \in \mathcal{M}$. However, h_a is smooth on $\mathcal{M}$ with $\mathrm{par}\left(\partial h(c)\right) = \{0\}$.

We end this section with a relationship between partial smoothness and the convergence analysis of Newton and quasi-Newton methods. Combining Proposition 2 and [24, Theorem 4.10], we have the following relationship between the sufficient conditions for superlinear convergence of the quasi-Newton method $\mathbf{Q}_k$ and the finite identification of an active manifold at a solution.

PROPOSITION 3 **(Finite Identification)**. *Let* $\mathcal{M}_{\overline{c}}$ *be as in* (34), *let Assumption 1 hold, and recall the notation of Definition 12. Let* $\overline{x} \in \mathrm{dom}\,(f)$ *and* $\overline{c} := c(\overline{x})$.
Suppose
(a) $\mathcal{M}_{\overline{c}}$ *satisfies the nondegeneracy condition,*
(b) the k*-strict complementarity condition of Definition 14 holds at* $(c, y) \in \mathbb{R}^m \times \mathbb{R}^m$,
(c) $M(\overline{x}) = \{\overline{y}\}$, *and*
(d) the second-order sufficient conditions of Theorem 2 are satisfied at $\overline{x}$.
Consider the neighborhood U *of* $(\overline{x}, \overline{y})$ *of Proposition 2, and a starting point* $(x^0, y^0) \in U$. *Suppose the sequence* $\left\{(x^k, y^k)\right\}_{k \in \mathbb{N}}$ *is generated from the optimality conditions for* $\mathbf{Q}_k$, *remains in* U *for all* $k \in \mathbb{N}$, *and satisfies* $(x^k, y^k) \neq (\overline{x}, \overline{y})$ *for all* $k \in \mathbb{N}$. *Then, the sufficient conditions for superlinear convergence of Proposition 2 imply* $c(x^k) + \nabla c(x^k)[x^{k+1} - x^k] \in \mathcal{M}_{\overline{c}}$ *for all large* k.

Proof. Since $x^k \to \overline{x}$, $d^k \to 0$. By continuity, $\widehat{c}_k := c(x^k) + \nabla c(x^k)[x^{k+1} - x^k] \to \overline{c}$. The quasi-Newton method (5) with $\mathbf{B}_k$ given by (31) implies $y^{k+1} \in \partial h(\widehat{c}_k)$, so $\{\widehat{c}_k\} \subset \operatorname{dom}(h)$. By Proposition 1, $h(\widehat{c}_k) \to h(\overline{c})$. Since $y^k \to \overline{y}$, $\operatorname{dist}\left(\overline{y} \,\middle|\, \partial h(\widehat{c}_k)\right) \le \left\|\overline{y} - y^{k+1}\right\| \to 0$. Then, by partial smoothness and [24, Theorem 4.10], $\widehat{c}_k \in \mathcal{M}_{\overline{c}}$ for all large k. □

7. Strong Metric Regularity and Local Quadratic Convergence of Newton's Method

The point of this section is to marry the partial smoothness hypothesis to the hypotheses used to establish strong metric subregularity in Section 6 to establish strong metric regularity of a solution mapping that is an appropriately defined local version of $g + G$ in (9). In addition, we establish the local quadratic convergence of the Newton method for $g + G$.

Definition 15 (Metric regularity). A set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *metrically regular* at $\overline{x}$ for $\overline{y}$ when $\overline{y} \in S(\overline{x})$, the graph of S is locally closed at $(\overline{x}, \overline{y})$, and there exists $\kappa \ge 0$ and neighborhoods U of $\overline{x}$ and V of $\overline{y}$ such that $\operatorname{dist}\left(x \,\middle|\, S^{-1}(y)\right) \le \kappa \operatorname{dist}\left(y \,\middle|\, S(x)\right)$ for all $(x, y) \in U \times V$. The infimum of κ over all (κ, U, V) satisfying the display is called the metric regularity modulus of S at $\overline{x}$ for $\overline{y}$, and is denoted $\operatorname{reg}(S; \overline{x}|\overline{y})$.

Definition 16 (Strong metric regularity). A set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *strongly metrically regular* at $\overline{x}$ for $\overline{y}$ when it is metrically regular at $\overline{x}$ for $\overline{y}$ and S^{-1} has a single-valued localization at $\overline{y}$ for $\overline{x}$. Equivalently, when S^{-1} has a Lipschitz continuous single-valued localization around $\overline{y}$ for $\overline{x}$.

7.1. Partly Smooth Problems

In this section, we make the following assumptions:

Assumption 2. Let h and f be as in **P**, with $\operatorname{dom}(h)$ given by the Rockafellar-Wets PLQ representation of Theorem 4. Let $(\overline{x}, \overline{y}) \in \operatorname{dom}(f) \times \mathbb{R}^m$, and set $\overline{c} := c(\overline{x})$, $\overline{k} := |\mathcal{K}(\overline{c})|$, where $\mathcal{K}(\overline{c}) = \left\{k_1, \dots, k_{\overline{k}}\right\}$ are the active indices given in Definition 7. Let $\mathcal{M}_{\overline{c}}$ be the active manifold defined in (34). By Lemma 6, there exists an integer ℓ such that $\ell = |I_k(\overline{c})|$ for any $k \in \mathcal{K}(\overline{c})$, and let $\overline{\mu}_j \in \mathbb{R}^\ell$ for $j \in \left\{1, \dots, \overline{k}\right\}$. With these specifications, we assume that

(a) c is $\mathcal{C}^3$-smooth,
(b) $\mathcal{M}_{\overline{c}}$ satisfies the nondegeneracy condition (in particular, $\overline{k} \ge 2$),
(c) f satisfies (SC) at $\overline{x}$ for $\overline{y}$; i.e., $\operatorname{Null}\left(\nabla c(\overline{x})^\top\right) \cap \operatorname{ri}\left(\partial h(\overline{c})\right) = \{\overline{y}\}$, so that in particular, as in (39), $J\overline{y} = \mathcal{Q}\overline{c} + \mathcal{B} + \widehat{\mathcal{A}}\overline{\mu}$, where $\overline{\mu} = (\overline{\mu}_1^\top, \dots, \overline{\mu}_{\overline{k}}^\top)^\top > 0$ by Lemma 10,
(d) $\overline{x}$ satisfies the second-order sufficient conditions of Theorem 2, i.e.,

$$h''(c(\overline{x}); \nabla c(\overline{x})d) + \left\langle d, \nabla^2_{xx} L(\overline{x}, \overline{y})d\right\rangle > 0 \quad \forall\, d \in \operatorname{Null}\left(A^\top \nabla c(\overline{x})\right) \setminus \{0\},$$

where, by Lemma 1, $M(\overline{x}) = \{\overline{y}\}$, and by Lemma 13, $D(\overline{x}) = \operatorname{Null}\left(A^\top \nabla c(\overline{x})\right)$.

The conditions (c) - (e) in Assumption 2 can be interpreted in terms of similar assumptions employed in classical NLP. Condition (c) corresponds to the linear independence of the active constraint gradients, (d) corresponds to strict complementary slackness, and (e) corresponds to the strong second-order sufficiency condition. The convergence results developed in this section subsume those known for NLP, since they follow from the case in which h is non finite-valued piecewise linear convex.

We begin with a key technical lemma important for establishing metric regularity.

Lemma 15. *In the notation of Definition 12, for any* $i, j \in \left\{1, \dots, \overline{k}\right\}$, $(Q_{k_i} - Q_{k_j})\operatorname{Null}\left(A^\top\right) \subset \operatorname{Ran}(A)$.

Proof. Let $w \in \mathrm{Null}\left(A^\top\right)$. By polyhedrality, there exists $|t| > 0$ such that $c_t := \overline{c} + tw \in \mathcal{M}_{\overline{c}}$. By Proposition 1, $\mathrm{dom}\,(\partial h) = \mathrm{dom}\,(h)$, so there exists $v \in \partial h(c_t)$ and $\overline{v} \in \partial h(\overline{c})$. By (39), $(v, \mu(c_t, v))$ and $(\overline{v}, \overline{\mu})$ satisfy $Jv = \mathcal{Q}c_t + \mathcal{B} + \widehat{\mathcal{A}}\mu(c_t, v)$ and $J\overline{v} = \mathcal{Q}\overline{c} + \mathcal{B} + \widehat{\mathcal{A}}\overline{\mu}$. Then for any $i, j \in \mathcal{K}(\overline{c})$,

$$
\begin{aligned}
0 &= (Q_{k_i} - Q_{k_j})c_t + A(P_i\mu(c_t, v)_i - P_j\mu(c_t, v)_j) + b_{k_i} - b_{k_j},\\
0 &= (Q_{k_i} - Q_{k_j})\overline{c} + A(P_i\overline{\mu}_i - P_j\overline{\mu}_j) + b_{k_i} - b_{k_j}.
\end{aligned}
$$

Subtracting the second equation from the first and rearranging gives

$$
(Q_{k_i} - Q_{k_j})w = t^{-1}A\left\{P_j(\mu(c_t, v)_j - \overline{\mu}_j) - P_i(\mu(c_t, v)_i - \overline{\mu}_i)\right\}. \tag{51}
$$

□

We now define a family of local approximations to $g + G$ for which strong metric regularity is established.

Definition 17 ($\mathcal{M}_{\overline{c}}$-restricted KKT Mappings). For a point $\overline{c} \in \mathcal{M}_{\overline{c}}$ and each $j \in \left\{1, \ldots, \overline{k}\right\}$, define $g_j : \mathbb{R}^{n+m+\ell} \to \mathbb{R}^{n+m+\ell+\ell}$ and G_0 by

$$
g_j(x, y, \mu_j) := \begin{pmatrix} \nabla c(x)^\top y \\ y - Q_{k_j}c(x) - b_{k_j} - AP_j\mu_j \\ A^\top[c(x) - \overline{c}] \\ -\mu_j \end{pmatrix} \text{ and } G_0 := \begin{pmatrix} \{0\}^n \\ \{0\}^m \\ \{0\}^\ell \\ \mathbb{R}^\ell_+. \end{pmatrix}
$$

and set $\overline{\boldsymbol{x}}_j := (\overline{x}, \overline{y}, \overline{\mu}_j) \in \mathbb{R}^{n+m+\ell}$, where $\overline{x}, \overline{y}, \overline{\mu}_j$ are as in Assumption 2. We call the mappings $g_j + G_0$ the $\mathcal{M}_{\overline{c}}$-restricted KKT Mappings.

Observe that

$$
\nabla g_j(x, y, \mu_j) = \begin{pmatrix} \nabla^2_{xx}L(x, y) & \nabla c(x)^\top & 0 \\ -Q_{k_j}\nabla c(x) & I & -AP_j \\ A^\top\nabla c(x) & 0 & 0 \\ 0 & 0 & -I_\ell \end{pmatrix} \text{ and } g_j(\overline{\boldsymbol{x}}_j) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -\overline{\mu}_j \end{pmatrix} \in -G_0 \text{ (see Assumption 2 (c))}.
$$

In parallel to the study in Section 5, we introduce the linearization of these mappings.

Definition 18 (Linearized $\mathcal{M}_{\overline{c}}$-restricted KKT Mappings). Let $\overline{c}$ and $\overline{k}$ be given by Assumption 2, and g_j and G_0 be as in Definition 17. For all $j \in \left\{1, \ldots, \overline{k}\right\}$, define the linearization of $g_j + G_0$ at $\boldsymbol{u} = (\widehat{x}, \widehat{y}, \widehat{\mu}_j)$ by

$$
\begin{aligned}
\mathcal{G}^j_{\boldsymbol{u}}(\boldsymbol{x}) &:= g_j(\boldsymbol{u}) + \nabla g_j(\boldsymbol{u})(\boldsymbol{x} - \boldsymbol{u}) + G_0, \text{ or equivalently,} \\
\mathcal{G}^j_{(\widehat{x}, \widehat{y}, \widehat{\mu}_j)}(x, y, \mu_j) &:= g_j(\widehat{x}, \widehat{y}, \widehat{\mu}_j) + \nabla g_j(\widehat{x}, \widehat{y}, \widehat{\mu}_j)\begin{pmatrix} x - \widehat{x} \\ y - \widehat{y} \\ \mu_j - \widehat{\mu}_j \end{pmatrix} + G_0.
\end{aligned} \tag{52}
$$

In the sequel, we will need to compute the normal cone to the graph of the linearized $\mathcal{M}_{\overline{c}}$-restricted KKT Mappings. For this purpose, we define for every $\boldsymbol{u} = (\widehat{x}, \widehat{y}, \widehat{\mu}_j)$ the function

$$
F_{\boldsymbol{u}}(\boldsymbol{x}, \boldsymbol{z}) := g_j(\boldsymbol{u}) + \nabla g_j(\boldsymbol{u})(\boldsymbol{x} - \boldsymbol{u}) - \boldsymbol{z} = \begin{pmatrix} \nabla c(\widehat{x})^\top y + \nabla^2_{xx}L(\widehat{x}, \widehat{y})[x - \widehat{x}] - z_1 \\ y - Q_{k_j}[c(\widehat{x}) + \nabla c(\widehat{x})[x - \widehat{x}]] - b_{k_j} - AP_j\mu_j - z_2 \\ A^\top[c(\widehat{x}) + \nabla c(\widehat{x})[x - \widehat{x}] - \overline{c}] - z_3 \\ -\mu_j - z_4 \end{pmatrix}. \tag{53}
$$

Then,

$$
\mathrm{gph}\,\mathcal{G}^j_{\boldsymbol{u}} = \left\{(\boldsymbol{x}, \boldsymbol{z}) \,\middle|\, F_{\boldsymbol{u}}(\boldsymbol{x}, \boldsymbol{z}) \in -G_0\right\}, \tag{54}
$$

with $\operatorname{dom}\left(\mathcal{G}^j_{\boldsymbol{u}}\right) = \mathbb{R}^{n+m+\ell}$. Explicitly,

$$\operatorname{gph}\mathcal{G}^j_{(\widehat{x},\widehat{y},\widehat{\mu}_j)} = \left\{ (x,y,\mu_j,z_1,z_2,z_3,z_4) \;\middle|\; \begin{array}{l} z_1 = \nabla c(\widehat{x})^\top y + \nabla^2_{xx} L(\widehat{x},\widehat{y})[x-\widehat{x}] \\ z_2 = y - Q_{k_j}[c(\widehat{x}) + \nabla c(\widehat{x})[x-\widehat{x}]] - b_{k_j} - AP_j\mu_j \\ z_3 = A^\top [c(\widehat{x}) + \nabla c(\widehat{x})[x-\widehat{x}] - \overline{c}] \\ z_4 \in -\mu_j + \mathbb{R}^\ell_+ \end{array} \right\}. \tag{55}$$

The next lemma shows that the error in the Newton iterates can be measured in terms of (x,y) alone, independent of the vectors μ_j. We omit its proof since it is a straightforward computation.

Lemma 16. *Let $\overline{x},\overline{y},\overline{\mu},\overline{c},\overline{k}$, and $\mathcal{Q}$ be as in Assumption 2, and g_j and G_0 be as in Definition 17. For any $j \in \left\{1,\dots,\overline{k}\right\}$, define $\eta_j : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n+m+\ell}$ by*

$$\eta_j(x,y) := \begin{pmatrix} \nabla c(x)^\top y \\ Q_{k_j}(\overline{c} - c(x)) \\ A^\top (c(x) - \overline{c}) \end{pmatrix}. \tag{56}$$

Observe that for any $(x,y,\mu_j) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^\ell$,

$$g_j(x,y,\mu_j) = \begin{pmatrix} \eta_j(x,y) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ y - \overline{y} + AP_j(\overline{\mu}_j - \mu_j) \\ 0 \\ -\mu_j \end{pmatrix} \text{ and } \nabla g_j(x,y,\mu_j) = \begin{pmatrix} \nabla\eta_j(x,y) & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & I & -AP_j \\ 0 & 0 & 0 \\ 0 & 0 & -I \end{pmatrix}$$

Set $\overline{\boldsymbol{x}}_j := (\overline{x},\overline{y},\overline{\mu}_j)$. Then, for any $\boldsymbol{u} := (\widehat{x},\widehat{y},\widehat{\mu}_j) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^\ell$,

$$\left\| F_{\boldsymbol{u}}(\overline{\boldsymbol{x}}_j, g_j(\overline{\boldsymbol{x}}_j)) \right\| = \left\| \eta_j(\widehat{x},\widehat{y}) + \nabla\eta_j(\widehat{x},\widehat{y}) \begin{pmatrix} \overline{x} - \widehat{x} \\ \overline{y} - \widehat{y} \end{pmatrix} - \eta_j(\overline{x},\overline{y}) \right\|, \tag{57}$$

since $\eta_j(\overline{x},\overline{y}) = 0$.

The following lemma uses the strict criticality assumption to show the normal cone to the graph of these linearizations is captured by the range of $\nabla F_{\overline{\boldsymbol{x}}_j}$.

Lemma 17. *Let $\overline{x},\overline{y},\overline{\mu},\overline{c},\overline{k}$, and $\mathcal{Q}$ be as in Assumption 2 and set $\overline{\boldsymbol{x}}_j := (\overline{x},\overline{y},\overline{\mu}_j)$. Then, for all $j \in \left\{1,\dots,\overline{k}\right\}$, the mapping $\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}$ in (54) has $N\left((\overline{\boldsymbol{x}}_j,\boldsymbol{0}) \mid \operatorname{gph}\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}\right) = \operatorname{Ran}(W)$, where*

$$W := \begin{pmatrix} \nabla^2_{xx} L(\overline{x},\overline{y}) & -\nabla c(\overline{x})^\top Q_{k_j} & \nabla c(\overline{x})^\top A \\ \nabla c(\overline{x}) & I_m & 0 \\ 0 & -P_j A^\top & 0 \\ -I_n & 0 & 0 \\ 0 & -I_m & 0 \\ 0 & 0 & -I_\ell \\ 0 & 0 & 0 \end{pmatrix}. \tag{58}$$

Proof. The set $\operatorname{gph}\mathcal{G}^j_{\overline{\boldsymbol{x}}_j} = \left\{(\boldsymbol{x},\boldsymbol{z}) \,\middle|\, F_{\overline{\boldsymbol{x}}_j}(\boldsymbol{x},\boldsymbol{z}) \in -G\right\}$ defined in (54) is closed with $(\overline{\boldsymbol{x}}_j,\mathbf{0}) \in \operatorname{gph}\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}$. In addition, $\overline{\mu}_j > 0$, $N\left(F_{\overline{\boldsymbol{x}}_j}(\overline{\boldsymbol{x}}_j,\mathbf{0}) \,|\, -G_0\right) = \mathbb{R}^{n+m+\ell} \times \{0\}^\ell$, and

$$\nabla F_{\overline{\boldsymbol{x}}_j}(\overline{\boldsymbol{x}}_j,\mathbf{0})^\top = \begin{pmatrix} \nabla^2_{xx}L(\overline{x},\overline{y}) & -\nabla c(\overline{x})^\top Q_{k_j} & \nabla c(\overline{x})^\top A & 0 \\ \nabla c(\overline{x}) & I_m & 0 & 0 \\ 0 & -P_jA^\top & 0 & I_\ell \\ -I_n & 0 & 0 & 0 \\ 0 & -I_m & 0 & 0 \\ 0 & 0 & -I_\ell & 0 \\ 0 & 0 & 0 & I_\ell \end{pmatrix} = \left(W \,|\, R\right),$$

where the matrix R is being defined by this expression. Combining the facts in the previous two sentences, the constraint qualification in [37, Theorem 6.14], for $N\left((\overline{\boldsymbol{x}}_j,\mathbf{0}) \,|\, \operatorname{gph}\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}\right)$ is the requirement that $\operatorname{Null}(W) = \{0\}$. If we verify $\operatorname{Null}(W) = \{0\}$, then $N\left((\overline{\boldsymbol{x}}_j,\mathbf{0}) \,|\, \operatorname{gph}\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}\right) = \operatorname{Ran}(W)$ by [37, Theorem 6.14]. But the presence of the identity matrices in W immediately give $\operatorname{Null}(W) = \{0\}$. □

The metric regularity of the mappings $g_j + G_0$ follow from the second-order sufficient conditions of Theorem 2.

LEMMA 18. *Let $\overline{x},\overline{y},\overline{\mu},\overline{c},\overline{k}$, and $\mathcal{Q}$ be as in Assumption 2, W as in (58) and set $\overline{\boldsymbol{x}}_j := (\overline{x},\overline{y},\overline{\mu}_j)$. For all $j \in \left\{1,\ldots,\overline{k}\right\}$,*

$$(\mathbf{0},-\boldsymbol{z}) \in N\left((\overline{\boldsymbol{x}}_j,\mathbf{0}) \,|\, \operatorname{gph}\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}\right) \iff \boldsymbol{z} = 0,$$

where $\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}$ is given by (54). Then, $\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}$ is metrically regular at $\overline{\boldsymbol{x}}_j$ for $\mathbf{0}$ and

$$\begin{pmatrix} \nabla^2_{xx}L(\overline{x},\overline{y}) & \nabla c(\overline{x})^\top & 0 \\ -Q_{k_j}\nabla c(\overline{x}) & I_m & -AP_j \\ A^\top\nabla c(\overline{x}) & 0 & 0 \end{pmatrix}$$

is nonsingular.

Proof. By Lemma 17, $N\left((\overline{\boldsymbol{x}}_j,\mathbf{0}) \,|\, \operatorname{gph}\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}\right) = \operatorname{Ran}(W)$, and so the statement

$$(\mathbf{0},-\boldsymbol{z}) \in N\left((\overline{\boldsymbol{x}}_j,\mathbf{0}) \,|\, \operatorname{gph}\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}\right) \iff \boldsymbol{z} = 0$$

is equivalent to

$$\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ -z_1 \\ -z_2 \\ -z_3 \\ -z_4 \end{pmatrix} = \begin{pmatrix} \nabla^2_{xx}L(\overline{x},\overline{y}) & -\nabla c(\overline{x})^\top Q_{k_j} & \nabla c(\overline{x})^\top A \\ \nabla c(\overline{x}) & I & 0 \\ 0 & -P_jA^\top & 0 \\ -I_n & 0 & 0 \\ 0 & -I_m & 0 \\ 0 & 0 & -I_\ell \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} d \\ v \\ w \end{pmatrix} \text{ for some } \begin{pmatrix} d \\ v \\ w \end{pmatrix}\right] \iff \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} = 0. \tag{59}$$

Since ($\Leftarrow$) is trivial, we only establish ($\Rightarrow$). Define $H := \nabla^2_{xx}L(\overline{x},\overline{y})$. Then the left-hand side of (59) becomes

$$0 = Hd - \nabla c(\overline{x})^\top Q_{k_j}v + \nabla c(\overline{x})^\top Aw, \tag{60}$$
$$0 = \nabla c(\overline{x})d + v, \tag{61}$$
$$0 = -P_jA^\top v, \tag{62}$$
$$z_1 = d,\ z_2 = v,\ z_3 = w,\ z_4 = 0.$$

Since $z_4 = 0$, we need only show $z_1 = z_2 = z_3 = 0$, which we establish by showing $d = v = w = 0$. First suppose $d \neq 0$. From (62) and Definition 12, $v \in \mathrm{Null}\left(A^\top\right)$. Then (61) and gives $\nabla c(\overline{x})d = -v \in \mathrm{Null}\left(A^\top\right)$. By Lemma 13, $d \in D(\overline{x}) \setminus \{0\}$. Taking the inner product on both sides of (60) with d and using (61) gives $d^\top H d = d^\top \nabla c(\overline{x})^\top Q_{k_j} v = -d^\top \nabla c(\overline{x})^\top Q_{k_j} \nabla c(\overline{x}) d$, so

$$d^\top \nabla c(\overline{x})^\top Q_{k_j} \nabla c(\overline{x}) d + d^\top H d = 0.$$

But the second-order sufficient conditions of Theorem 2 imply that for any $j \in \left\{1, \ldots, \overline{k}\right\}$,

$$d^\top \nabla c(\overline{x})^\top Q_{k_j} \nabla c(\overline{x}) d + d^\top H d > 0.$$

This contradiction implies $d = 0$. But then $v = 0$ by (61). Finally, (60) states that w must satisfy $Aw \in \mathrm{Null}\left(\nabla c(\overline{x})^\top\right) \cap \mathrm{Ran}\,(A) = \{0\}$. By the nondegeneracy condition of Definition 13, $w = 0$. Equation (54) gives local closedness of $\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}$ at $(\overline{\boldsymbol{x}}_j, \mathbf{0})$, so the coderivative criterion for metric regularity [13, Theorem 4C.2] implies $\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}$ is metrically regular at $\overline{\boldsymbol{x}}_j$ for $\mathbf{0}$, as required. □

The metric regularity of the mappings $\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}$ imply a parameterized uniform version of metric regularity, where we allow $\overline{\boldsymbol{x}}_j$ to move.

Lemma 19. *Let $\overline{x}, \overline{y}, \overline{\mu}, \overline{c}, \overline{k}$, and $\mathcal{Q}$ be as in Assumption 2, set $\overline{\boldsymbol{x}}_j := (\overline{x}, \overline{y}, \overline{\mu}_j)$, and let $\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}$ be given by (54). For all $j \in \left\{1, \ldots, \overline{k}\right\}$, there exists a neighborhood $U_j \subset \mathbb{R}^{n+m+\ell}$ of $\overline{\boldsymbol{x}}_j$ and a neighborhood $V_j \subset \mathbb{R}^{n+m+\ell+\ell}$ of $\mathbf{0}$ such that the mapping*

$$(\boldsymbol{u}, \boldsymbol{z}) \mapsto \mathcal{G}^{-j}_{\boldsymbol{u}}(\boldsymbol{z}) := \left(\mathcal{G}^j_{\boldsymbol{u}}\right)^{-1}(\boldsymbol{z}) \text{ for } (\boldsymbol{u}, \boldsymbol{z}) \in U_j \times V_j$$

is single-valued with $\mathcal{G}^{-j}_{\boldsymbol{u}}(\mathbf{0}) \in U_j$.

Proof. Fix $j \in \left\{1, \ldots, \overline{k}\right\}$. By Lemma 18 and [13, Theorem 6D.1], for every $\lambda > \mathrm{reg}(\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}; \overline{\boldsymbol{x}}_j | \mathbf{0})$ there exists $a > 0$ and $b > 0$ such that

$$\mathrm{dist}\left(\boldsymbol{x} \;\middle|\; \mathcal{G}^{-j}_{\boldsymbol{u}}(\boldsymbol{z})\right) \leq \lambda \mathrm{dist}\left(\boldsymbol{z} \;\middle|\; \mathcal{G}^j_{\boldsymbol{u}}(\boldsymbol{x})\right), \quad \text{for every } \boldsymbol{u}, \boldsymbol{x} \in \overline{\boldsymbol{x}}_j + a\mathbb{B}, \boldsymbol{z} \in b\mathbb{B}. \tag{63}$$

By reducing a, if necessary, we may assume the conclusion of Lemma 14 holds on $\overline{\boldsymbol{x}}_j + a\mathbb{B}$. We follow the argument given in [13, Theorem 6D.2] by recalling (56) and choosing

$$L > \mathrm{lip}(\nabla \eta_j; (\overline{x}, \overline{y})) := \limsup_{\substack{(x,y),(x',y') \to (\overline{x},\overline{y}) \\ (x,y) \neq (x',y')}} \frac{\left\|\nabla \eta_j(x,y) - \nabla \eta_j(x',y')\right\|}{\left\|(x,y) - (x',y')\right\|}, \text{ and } \gamma > \frac{1}{2}\lambda L.$$

Define $\overline{a} := \min\left\{\frac{1}{\gamma}, a\right\} > 0$, $U_j := \overline{\boldsymbol{x}}_j + \overline{a}\mathbb{B}$, and $V_j := b\mathbb{B}$. We first establish nonemptiness of $\mathcal{G}^{-j}_{\boldsymbol{u}}(\boldsymbol{z})$. Fix $\boldsymbol{x} = \overline{\boldsymbol{x}}_j$, and choose any $(\boldsymbol{u}, \boldsymbol{z}) \in U_j \times V_j$, and consider two cases in (63). If $\mathrm{dist}\left(\boldsymbol{z} \;\middle|\; \mathcal{G}^j_{\boldsymbol{u}}(\overline{\boldsymbol{x}}_j)\right) = 0$, then by closedness of the set $\mathcal{G}^j_{\boldsymbol{u}}(\overline{\boldsymbol{x}}_j)$, it follows that $\overline{\boldsymbol{x}}_j \in \mathcal{G}^{-j}_{\boldsymbol{u}}(\boldsymbol{z})$. On the other hand, if $0 < \mathrm{dist}\left(\boldsymbol{z} \;\middle|\; \mathcal{G}^j_{\boldsymbol{u}}(\overline{\boldsymbol{x}}_j)\right) < \infty$, where finiteness is guaranteed because $\mathrm{dom}\left(\mathcal{G}^j_{\boldsymbol{u}}\right) = \mathbb{R}^{m+n+\ell}$. Then the implication

$$\mathrm{dist}\left(\overline{\boldsymbol{x}}_j \;\middle|\; \mathcal{G}^{-j}_{\boldsymbol{u}}(z)\right) \leq \lambda \mathrm{dist}\left(\boldsymbol{z} \;\middle|\; \mathcal{G}^j_{\boldsymbol{u}}(\overline{\boldsymbol{x}}_j)\right) \Longrightarrow \mathrm{dist}\left(\overline{\boldsymbol{x}}_j \;\middle|\; \mathcal{G}^{-j}_{\boldsymbol{u}}(z)\right) < \infty$$

holds, so in both cases $\mathcal{G}^{-j}_{\boldsymbol{u}}(z) \neq \emptyset$.

We now show single-valuedness. For the same j, $\boldsymbol{u}$, and $\boldsymbol{z}$, write $\boldsymbol{u} = (\widehat{x}, \widehat{y}, \widehat{\mu}_j)$, and suppose there are two points $\boldsymbol{x}_1 = (x_1, y_1, \mu_{j_1})$, $\boldsymbol{x}_2 = (x_2, y_2, \mu_{j_2})$ satisfying $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{G}_{\boldsymbol{u}}^{-j}(\boldsymbol{z})$. Then subtracting the equations in (55) gives

$$0 = \nabla^2_{xx} L(\widehat{x}, \widehat{y})[x_2 - x_1] + \nabla c(\widehat{x})^\top (y_2 - y_1) \tag{64}$$
$$y_2 - y_1 = Q_{k_j} \nabla c(\widehat{x})[x_2 - x_1] + AP_j(\mu_{j_2} - \mu_{j_1}) \tag{65}$$
$$0 = A^\top \nabla c(\widehat{x})[x_2 - x_1]. \tag{66}$$

Then $\nabla c(\widehat{x})[x_2 - x_1] \in \mathrm{Null}\left(A^\top\right)$. Suppose $x_2 \neq x_1$. Taking the inner product on both sides of (64) and using the choice of $\overline{a}$ in accordance with Lemma 14,

$$\begin{aligned}
0 &= [x_2 - x_1]^\top \nabla^2_{xx} L(\widehat{x}, \widehat{y})[x_2 - x_1] + [x_2 - x_1]^\top \nabla c(\widehat{x})^\top (y_2 - y_1) && \text{by (64)}\\
&= [x_2 - x_1]^\top \nabla^2_{xx} L(\widehat{x}, \widehat{y})[x_2 - x_1] + [x_2 - x_1]^\top \nabla c(\widehat{x})^\top [Q_{k_j} \nabla c(\widehat{x})[x_2 - x_1] + AP_j(\mu_{j_2} - \mu_{j_1})] && \text{by (65)}\\
&= [x_2 - x_1]^\top \nabla^2_{xx} L(\widehat{x}, \widehat{y})[x_2 - x_1] + [x_2 - x_1]^\top \nabla c(\widehat{x})^\top Q_{k_j} \nabla c(\widehat{x})[x_2 - x_1] && \text{by (66)}\\
&> 0,
\end{aligned}$$

so $x_2 = x_1$. But then (64), (65), and Lemma 14 imply

$$y_2 - y_1 \in \mathrm{Null}\left(\nabla c(\widehat{x})^\top\right) \cap \mathrm{Ran}\,(A) = \{0\}\,,$$

so $y_2 = y_1$. The nondegeneracy condition of Definition 13 and (65) together imply

$$0 = AP_j(\mu_{j_2} - \mu_{j_1}) \Longrightarrow \mu_{j_2} = \mu_{j_1},$$

so single-valuedness is established. We conclude the proof by following the proof given in [13, Theorem 6D.2] and write $(x, y, \mu_j) = \boldsymbol{x} = \mathcal{G}_{\boldsymbol{u}}^{-j}(0)$. Then the quadratic bound lemma and the choice of γ gives

$$\begin{aligned}
\left\| \begin{pmatrix} x - \overline{x} \\ y - \overline{y} \end{pmatrix} \right\| &\le \left\| \boldsymbol{x} - \overline{\boldsymbol{x}}_j \right\| \\
&= \mathrm{dist}\left(\overline{\boldsymbol{x}}_j \,\middle|\, \mathcal{G}_{\boldsymbol{u}}^{-j}(0) \right) \\
&\le \lambda \mathrm{dist}\left(\boldsymbol{0} \,\middle|\, \mathcal{G}_{\boldsymbol{u}}^{j}(\overline{\boldsymbol{x}}_j) \right) \\
&\le \frac{2\gamma}{L} \mathrm{dist}\left(\boldsymbol{0} \,\middle|\, \mathcal{G}_{\boldsymbol{u}}^{j}(\overline{\boldsymbol{x}}_j) \right) \\
&\le \frac{2\gamma}{L} \left\| g_j(\boldsymbol{u}) + \nabla g_j(\boldsymbol{u})(\overline{\boldsymbol{x}}_j - \boldsymbol{u}) - g_j(\overline{\boldsymbol{x}}_j) \right\| && \text{by (52) and } -g_j(\overline{\boldsymbol{x}}_j) \in G_0 \\
&= \frac{2\gamma}{L} \left\| F_{\boldsymbol{u}}(\overline{\boldsymbol{x}}_j, g_j(\overline{\boldsymbol{x}}_j)) \right\| && \text{by (53)} \\
&= \frac{2\gamma}{L} \left\| \eta_j(\widehat{x}, \widehat{y}) + \nabla \eta_j(\widehat{x}, \widehat{y}) \begin{pmatrix} \overline{x} - \widehat{x} \\ \overline{y} - \widehat{y} \end{pmatrix} - \eta_j(\overline{x}, \overline{y}) \right\| && \text{by (57)} \\
&\le \gamma \left\| \begin{pmatrix} \widehat{x} - \overline{x} \\ \widehat{y} - \overline{y} \end{pmatrix} \right\|^2 \\
&\le \gamma \left\| \boldsymbol{u} - \overline{\boldsymbol{x}}_j \right\|^2 \\
&< \overline{a},
\end{aligned}$$

so $\boldsymbol{x} = \mathcal{G}_{\boldsymbol{u}}^{j}(\boldsymbol{0}) \in U_j$. □

Our work so far implies that Newton's method applied to the individual mappings $\mathcal{G}_{\overline{\boldsymbol{x}}_j}^{j}$ exhibit local quadratic convergence.

THEOREM 6. *Let $\overline{x}, \overline{y}, \overline{\mu}, \overline{c}, \overline{k}$, and $\mathcal{Q}$ be as in Assumption 2, set $\overline{\boldsymbol{x}}_j := (\overline{x}, \overline{y}, \overline{\mu}_j)$, and let $\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}$ be given by (54). Then, the mappings $\left\{\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}\right\}_{j=1}^{\overline{k}}$ are strongly metrically regular (see Definition 16) at $\overline{\boldsymbol{x}}_j$ for $\mathbf{0}$. Moreover, for all $j \in \left\{1, \ldots, \overline{k}\right\}$, there exists a neighborhood U_j of $\overline{\boldsymbol{x}}_j$ such that, for every $\boldsymbol{x}^0 \in U_j$, there is a unique sequence $\boldsymbol{x}^k = (x^k, y^k, \mu^k_j) \subset U_j$ generated by Newton's method for $g_j + G_0$ (4). Both this sequence, and the sequence (x^k, y^k), converge at a quadratic rate to $\overline{\boldsymbol{x}}_j$ and $(\overline{x}, \overline{y})$ respectively.*

Proof. The metric regularity at $\overline{\boldsymbol{x}}_j$ for $\mathbf{0}$ was established in Lemma 18. Lemma 19 with $u = \overline{\boldsymbol{x}}_j$ shows $\mathcal{G}^{-j}_{\overline{\boldsymbol{x}}_j}$ has a single-valued localization around $\mathbf{0}$ for $\overline{\boldsymbol{x}}_j$, so the strong metric regularity of $\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}$ at $\overline{\boldsymbol{x}}_j$ for $\mathbf{0}$ follows.

For the second claim, we again follow the proof in [13, Theorem 6D.2] by taking U_j as in Lemma 19, and choosing any $\boldsymbol{x}^0 \in U_j$. Following the proof of the final claim of Lemma 19, we find, for every $k \geq 1$, the existence and uniqueness of $\boldsymbol{x}^k$ given $\boldsymbol{x}^{k-1}$ satisfying

$$\mathbf{0} \in \mathcal{G}^j_{\boldsymbol{x}^{k-1}}(\boldsymbol{x}^k), \ \left\|\begin{pmatrix} x^k - \overline{x} \\ y^k - \overline{y} \end{pmatrix}\right\| \leq \left\|\boldsymbol{x}^k - \overline{\boldsymbol{x}}_j\right\| \leq \gamma \left\|\begin{pmatrix} x^{k-1} - \overline{x} \\ y^{k-1} - \overline{y} \end{pmatrix}\right\|^2 \leq \gamma \left\|\boldsymbol{x}^{k-1} - \overline{\boldsymbol{x}}_j\right\|^2, \text{ and } \boldsymbol{x}^k \in U_j.$$

Moreover, since $\theta := \gamma\left\|\boldsymbol{x}^0 - \overline{\boldsymbol{x}}_j\right\| < \gamma \overline{a} < 1$, $\left\|\boldsymbol{x}^k - \overline{\boldsymbol{x}}_j\right\| \leq \theta^{2^k - 1}\left\|\boldsymbol{x}^0 - \overline{\boldsymbol{x}}_j\right\|^2$ for all $k \geq 1$, which completes the proof of quadratic convergence of both sequences. □

We now move from an isolated analysis of the mappings $\mathcal{G}^j_{\boldsymbol{u}}$ to how they behave as a whole. The goal is to guarantee the y obtained by solving $\mathbf{0} \in \mathcal{G}^j_{\boldsymbol{u}}(\boldsymbol{x})$ at some $\boldsymbol{u} = (\widehat{x}, \widehat{y}, \widehat{\mu}_j)$ for $\boldsymbol{x} = (x, y, \mu_j)$ has $y \in \partial h(c(\widehat{x}) + \nabla c(\widehat{x})[x - \widehat{x}])$.

THEOREM 7. *Let $\overline{x}, \overline{y}, \overline{\mu}, \overline{c}, \overline{k}$, and $\mathcal{Q}$ be as in Assumption 2, set $\overline{\boldsymbol{x}}_j := (\overline{x}, \overline{y}, \overline{\mu}_j)$, and let $\mathcal{G}^j_{\overline{\boldsymbol{x}}_j}$ be given by (54). Suppose $i \neq j$ and $i, j \in \left\{1, \ldots, \overline{k}\right\}$. There exists a neighborhood $\mathcal{N}$ of $(\overline{x}, \overline{y}, \overline{\mu}_1, \ldots, \overline{\mu}_{\overline{k}}) =: (\overline{x}, \overline{y}, \overline{\mu}) \in \mathbb{R}^{n+m+\overline{k}\ell}$ such that, if $(\widehat{x}, \widehat{y}, \widehat{\mu}_1, \ldots, \widehat{\mu}_{\overline{k}}) \in \mathcal{N}$ and $\boldsymbol{u}_j := (\widehat{x}, \widehat{y}, \widehat{\mu}_j)$, $\boldsymbol{u}_i := (\widehat{x}, \widehat{y}, \widehat{\mu}_i)$, with $\widehat{\mu}_i > 0$ and $\widehat{\mu}_j > 0$, then*

$$\boldsymbol{x}_j := \mathcal{G}^{-j}_{\boldsymbol{u}_j}(\mathbf{0}) = \begin{pmatrix} x_j \\ y_j \\ \mu_j \end{pmatrix}, \quad \boldsymbol{x}_i := \mathcal{G}^{-i}_{\boldsymbol{u}_i}(\mathbf{0}) = \begin{pmatrix} x_i \\ y_i \\ \mu_i \end{pmatrix} \ \textit{satisfy} \ \begin{pmatrix} x_j \\ y_j \end{pmatrix} = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \ \textit{for all } i, j \in \left\{1, \ldots, \overline{k}\right\}. \tag{67}$$

That is, there exists $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $(x, y) = (x_i, y_i)$ for all $i \in \left\{1, \ldots, \overline{k}\right\}$. Moreover,

(i) $c(\widehat{x}) + \nabla c(x)[x - \widehat{x}] \in \mathcal{M}_{\overline{c}}$,

(ii) $\mu(c(\widehat{x}) + \nabla c(\widehat{x})[x - \widehat{x}], y)_j = \mu_j > 0$ *for all* $j \in \left\{1, \ldots, \overline{k}\right\}$,

(iii) $y \in \operatorname{ri}\left(\partial h(c(\widehat{x}) + \nabla c(\widehat{x})[x - \widehat{x}])\right)$,

where the mapping $\mu(c, y)$ is defined in Lemma 9.

Proof. For $j \in \left\{1, \ldots, \overline{k}\right\}$, define $\pi_j : \mathbb{R}^{n+m+\overline{k}\ell} \to \mathbb{R}^{n+m+\ell}$ by $\pi_j(x, y, \mu_1, \ldots, \mu_j, \ldots, \mu_{\overline{k}}) := (x, y, \mu_j)$. We first show there exists a neighborhood $\mathcal{N}$ of $(\overline{x}, \overline{y}, \overline{\mu}_1, \ldots, \overline{\mu}_{\overline{k}})$ such that, for all $j \in \left\{1, \ldots, \overline{k}\right\}$ and all $(\widehat{x}_j, \widehat{y}_j, \widehat{\mu}_j) = \boldsymbol{u}_j \in \mathcal{N}_j := \pi_j(\mathcal{N})$,

(a) the mappings $\left\{\mathcal{G}^{-j}_{\boldsymbol{u}_j}(\mathbf{0})\right\}_{j=1}^{\overline{k}}$ are single-valued with $\mathcal{G}^{-j}_{\boldsymbol{u}_j}(\mathbf{0}) \in \mathcal{N}_j$,

(b) μ_j associated to $\mathcal{G}^{-j}_{\boldsymbol{u}_j}(\mathbf{0})$ has $\mu_j > 0$,

(c) the condition (50) is satisfied at all $(x, y, \mu_j) \in \mathcal{N}_j$, and

(d) $c(\widehat{x}_j) + \nabla c(\widehat{x}_j)[x_j - \widehat{x}_j] \in \mathcal{M}_{\overline{c}}$, where $(x_j, y_j, \mu_j) = \mathcal{G}^{-j}_{(\widehat{x}, \widehat{y}, \widehat{\mu}_j)}(\mathbf{0})$.

Parts (a), (b), and (c) are a consequence of Lemma 19. We now justify (d). For any $j \in \left\{1,\ldots,\overline{k}\right\}$, the definition of $(x_j, y_j, \mu_j) = \mathcal{G}^{-j}_{(\widehat{x},\widehat{y},\widehat{\mu}_j)}(\mathbf{0})$ implies, in particular, $A^\top[c(\widehat{x}_j)+\nabla c(\widehat{x}_j)[x_j-\widehat{x}_j]-c(\overline{x})] = 0$. By the polyhedral structure of $\mathcal{M}_{\overline{c}}$, for any $w \in \mathrm{Null}\left(A^\top\right) = T\left(\overline{c}\,|\,\mathcal{M}_{\overline{c}}\right)$, there exists $\tau > 0$ such that $\overline{c}+tw \in \mathcal{M}_{\overline{c}}$ for all $|t| < \tau$. Lemma 19 argued that, for all sufficiently small $\epsilon > 0$,

$$\mathcal{G}^{-j}_{\boldsymbol{u}}(\mathbf{0}) \in (\overline{\boldsymbol{x}}_j + \epsilon\mathbb{B}) \text{ for all } \boldsymbol{u} \in \overline{\boldsymbol{x}}_j + \epsilon\mathbb{B} \text{ (see (a))}. \tag{68}$$

The continuity of c and (68) imply that for $\boldsymbol{u}_j$ sufficiently close to $\overline{\boldsymbol{x}}_j$, $c(\widehat{x}_j)+\nabla c(\widehat{x}_j)[x_j-\widehat{x}_j]$ can be made as close to $c(\overline{x})$ as desired. Then there exists a neighborhood of $(\overline{x},\overline{y},\overline{\mu}_j)$ such (d) holds. The neighborhood $\mathcal{N}$ also exists because there are only finitely many indices j in consideration.

Now let $\boldsymbol{u}_j := (\widehat{x},\widehat{y},\widehat{\mu}_j) \in \mathcal{N}_j$, $\boldsymbol{u}_i := (\widehat{x},\widehat{y},\widehat{\mu}_i) \in \mathcal{N}_i$, with $\widehat{\mu}_i > 0$ and $\widehat{\mu}_j > 0$, and denote

$$\mathcal{G}^{-j}_{\boldsymbol{u}_j}(\mathbf{0}) = \begin{pmatrix} x_j \\ y_j \\ \mu_j \end{pmatrix}, \quad \mathcal{G}^{-i}_{\boldsymbol{u}_i}(\mathbf{0}) = \begin{pmatrix} x_i \\ y_i \\ \mu_i \end{pmatrix}.$$

By (55),

$$\begin{aligned}
0 &= \nabla^2_{xx} L(\widehat{x},\widehat{y})[x_j - x_i] + \nabla c(\widehat{x})^\top (y_j - y_i) && (69)\\
y_i &= Q_{k_i}(c(\widehat{x}) + \nabla c(\widehat{x})[x_i - \widehat{x}]) + AP_i\mu_i + b_{k_i} && (70)\\
y_j &= Q_{k_j}(c(\widehat{x}) + \nabla c(\widehat{x})[x_j - \widehat{x}]) + AP_j\mu_j + b_{k_j} && (71)\\
0 &= A^\top \nabla c(\widehat{x})[x_j - x_i] && (72)
\end{aligned}$$

Define $\widehat{c}_i := c(\widehat{x}) + \nabla c(\widehat{x})[x_i - \widehat{x}] \in \mathcal{M}_{\overline{c}}$ by (d). By Assumption 3, $\overline{y} = Q_{k_i}\overline{c} + b_{k_i} + AP_i\overline{\mu}_i = Q_{k_j}\overline{c} + b_{k_j} + AP_j\overline{\mu}_j$, and in particular,

$$Q_{k_i}\overline{c} + b_{k_i} - b_{k_j} = Q_{k_j}\overline{c} + AP_j\overline{\mu}_j - AP_i\overline{\mu}_i. \tag{73}$$

Then (51) with $w := \widehat{c}_i - \overline{c} \in \mathrm{Null}\left(A^\top\right)$, $t = 1$, and any $y \in \partial h(\widehat{c}_i)$ gives

$$\begin{aligned}
y_i &= Q_{k_i}w + Q_{k_i}\overline{c} + b_{k_i} + AP_i\mu_i\\
&= \left(Q_{k_j}w + A\left\{P_j(\mu(\widehat{c}_i,y)_j - \overline{\mu}_j) - P_i(\mu(\widehat{c}_i,y)_i - \overline{\mu}_i)\right\}\right) + Q_{k_i}\overline{c} + b_{k_i} + AP_i\mu_i + b_{k_j} - b_{k_j}\\
&= Q_{k_j}w + b_{k_j} + [Q_{k_i}\overline{c} + b_{k_i} - b_{k_j}] + AP_i\mu_i + A\left\{P_j(\mu(\widehat{c}_i,y)_j - \overline{\mu}_j) - P_i(\mu(\widehat{c}_i,y)_i - \overline{\mu}_i)\right\}\\
&= Q_{k_j}[\widehat{c}_i - \overline{c}] + b_{k_j} + [Q_{k_j}\overline{c} + AP_j\overline{\mu}_j - AP_i\overline{\mu}_i] + AP_i\mu_i + A\left\{P_j(\mu(\widehat{c}_i,y)_j - \overline{\mu}_j) - P_i(\mu(\widehat{c}_i,y)_i - \overline{\mu}_i)\right\}\\
&= Q_{k_j}\widehat{c}_i + b_{k_j} + AP_i[\mu_i - \mu(\widehat{c}_i,y)_i] + AP_j\mu(\widehat{c}_i,y)_j\\
&\in y_j + Q_{k_j}\nabla c(\widehat{x})[x_i - x_j] + \mathrm{Ran}\,(A)
\end{aligned}$$

where the fourth equivalence follows from (73). This implies

$$y_j - y_i - Q_{k_j}\nabla c(\widehat{x})[x_j - x_i] \in \mathrm{Ran}\,(A)\,. \tag{74}$$

Taking the inner product on both sides of (69) with $x_j - x_i$ gives

$$\begin{aligned}
0 &= [x_j - x_i]^\top \nabla^2_{xx} L(\widehat{x},\widehat{y})[x_j - x_i] + [x_j - x_i]^\top \nabla c(\widehat{x})^\top (y_j - y_i)\\
&= [x_j - x_i]^\top \nabla^2_{xx} L(\widehat{x},\widehat{y})[x_j - x_i] + [x_j - x_i]^\top \nabla c(\widehat{x})^\top Q_{k_j} \nabla c(\widehat{x})[x_j - x_i] \text{ by (74), (72).}
\end{aligned}$$

By Lemma 14 and (72), $x_i = x_j$. Then (74), (69), and (c) imply $y_i - y_j \in \mathrm{Ran}\,(A) \cap \mathrm{Null}\left(\nabla c(\widehat{x})^\top\right) = \{0\}$, which proves (67).

Since i and j were arbitrary, letting x and y denote the common values of the first two components of $\mathcal{G}_{\boldsymbol{u}_j}^{-j}(\mathbf{0})$ for each $j \in \left\{1,\ldots,\overline{k}\right\}$. Then $Jy = \mathcal{Q}(c(\widehat{x})+\nabla c(\widehat{x})[x-\widehat{x}])+\mathcal{B}+\widehat{\mathcal{A}}\mu$, with $c(\widehat{x})+\nabla c(\widehat{x})[x-\widehat{x}]\in\mathcal{M}_{\overline{c}}$, and $\mu_1,\ldots,\mu_{\overline{k}}>0$. By (39) and Lemma 10, $\mu(c(\widehat{x})+\nabla c(\widehat{x})[x-\widehat{x}],y)_j=\mu_j>0$, with $y\in\operatorname{ri}\left(\partial h(c(\widehat{x})+\nabla c(\widehat{x})[x-\widehat{x}])\right)$. □

Our final theorem integrates the ideas from Section 6 and our work in this section to establish the local quadratic convergence of Newton's method for **P**.

THEOREM 8. *Let $\overline{x},\overline{y},\overline{\mu},\overline{c},\overline{k}$, and $\mathcal{Q}$ be as in Assumption 2, set $\overline{\boldsymbol{x}}_j := (\overline{x},\overline{y},\overline{\mu}_j)$, and let $\mathcal{G}_{\overline{\boldsymbol{x}}_j}^{j}$ be given by (54). There exists a neighborhood $\mathcal{N}$ of $(\overline{x},\overline{y},\overline{\mu})$ on which the conclusions of Lemma 14 are satisfied such that if $(x^0,y^0,\mu^0)\in\mathcal{N}$, then there exists a unique sequence $\left\{(x^k,y^k,\mu^k)\right\}_{k\in\mathbb{N}}$ satisfying the optimality conditions of $\mathbf{P}_k$ for all $k\in\mathbb{N}$, with*
(a) $c(x^{k-1})+\nabla c(x^{k-1})[x^k-x^{k-1}]\in\mathcal{M}_{\overline{c}}$,
(b) $\mu(c(x^{k-1})+\nabla c(x^{k-1})[x^k-x^{k-1}],y^k)_j>0$ for all $j\in\left\{1,\ldots,\overline{k}\right\}$,
(c) $y^k\in\operatorname{ri}\left(\partial h(c(x^{k-1})+\nabla c(x^{k-1})[x^k-x^{k-1}])\right)$,
(d) $H_{k-1}[x^k-x^{k-1}]+\nabla c(x^{k-1})^\top y^k=0$,
(e) x^k-x^{k-1} is a strong local minimizer of $d\mapsto h(c(x^{k-1})+\nabla c(x^{k-1})d)+\frac{1}{2}d^\top H_{k-1}d$.
Moreover, the sequence (x^k,y^k) converges to $(\overline{x},\overline{y})$ at a quadratic rate.

Proof. All claims except (e) follow from Theorem 6 and Theorem 7. By Lemma 22 and Lemma 23, claim (e) is equivalent to showing

$$h''(c(x^{k-1})+\nabla c(x^{k-1})[x^k-x^{k-1}];\nabla c(x^{k-1})\delta)+\delta^\top H_{k-1}\delta>0\ \forall\,\delta\in\operatorname{Null}\left(A^\top\nabla c(x^{k-1})\right)\setminus\{0\}. \quad (75)$$

Using (15) and partial smoothness,

$$h''(c(x^{k-1})+\nabla c(x^{k-1})[x^k-x^{k-1}];\nabla c(x^{k-1})\delta)=\delta^\top\nabla c(x^{k-1})^\top Q_j\nabla c(x^{k-1})\delta,\quad \forall\, j\in\mathcal{K}(\overline{c}),$$

so (50) gives (75). □

REMARK 10. The fact that $\left\{x^k-x^{k-1}\right\}$ are strong local minimizers does not mean that there are not other critical points outside the neighborhood of quadratic convergence. It may be that at any iteration the problem $\widehat{\mathbf{P}}$ does not have a finite optimal value, in particular, should there exist directions of negative curvature orthogonal to the manifold.

7.2. Smooth Problems

In this section, we make the following assumptions:

ASSUMPTION 3. Let h and f be as in **P**, with $\operatorname{dom}(h)$ given by the Rockafellar-Wets PLQ representation of Theorem 4. Let $(\overline{x},\overline{y})\in\operatorname{dom}(f)\times\mathbb{R}^m$, and set $\overline{c}:=c(\overline{x})$, $\overline{k}=|\mathcal{K}(\overline{c})|$, where $\mathcal{K}(\overline{c})$ are the active indices given in Definition 7. Let $\mathcal{M}_{\overline{c}}$ be the active manifold defined in (34). We assume that
(a) c is $\mathcal{C}^3$-smooth,
(b) $\mathcal{K}(\overline{c})=\{k_0\}$,
(c) $\overline{x}$ satisfies the second-order sufficient conditions of Theorem 2,

REMARK 11. Since $\overline{k}=1$, we omit reference to the index k_0 for the rest of this section.

REMARK 12. By (b), $c(\overline{x})\in\operatorname{int}\operatorname{dom}(h)$ and $\partial h(\overline{c})=\{\overline{y}\}$. Then, (c) becomes

$$\overline{y}=Q\overline{c}+b,\ \nabla c(\overline{x})^\top\overline{y}=0,\ d^\top\nabla c(\overline{x})^\top Q\nabla c(\overline{x})d+d^\top\nabla^2_{xx}L(\overline{x},\overline{y})d>0\ \forall\, d\in\mathbb{R}^n\setminus\{0\},\text{ where } D(\overline{x})=\mathbb{R}^n.$$

As in Lemma 14, we have the following stability result.

LEMMA 20. *Suppose $d^\top\nabla c(\overline{x})^\top Q\nabla c(\overline{x})d+d^\top\nabla^2_{xx}L(\overline{x},\overline{y})d>0$ for all $d\in\mathbb{R}^n\setminus\{0\}$. Then, there exists a neighborhood $\mathcal{N}$ of $(\overline{x},\overline{y})$ such that if $(x,y)\in\mathcal{N}$ then,*

$$d^\top\nabla c(x)^\top Q\nabla c(x)d+d^\top\nabla^2_{xx}L(x,y)d>0,\quad \forall\ d\in\mathbb{R}^n\setminus\{0\}, \quad (76)$$

and $c(x)\in\operatorname{int}\operatorname{dom}(h)$.

The differentiability of h at $c(\overline{x})$ suggests the following KKT mapping (9).

DEFINITION 19 ($\mathcal{M}_{\overline{c}}$-RESTRICTED SMOOTH KKT MAPPING). Define $g:\mathbb{R}^{n+m}\to\mathbb{R}^{n+m}$ by

$$g(x,y) := \begin{pmatrix} \nabla c(x)^\top y \\ y - Qc(x) - b \end{pmatrix}, \quad G := \{0\}^{n+m},$$

and set $\overline{\boldsymbol{x}} := (\overline{x},\overline{y})$, so that,

$$\nabla g(x,y) = \begin{pmatrix} \nabla^2_{xx} L(x,y) & \nabla c(x)^\top \\ -Q\nabla c(x) & I_m \end{pmatrix}, \quad g(\overline{x},\overline{y}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Assumption (c) implies $\nabla g(\overline{x},\overline{y})$ is nonsingular. Consequently, the Newton method (4) corresponds to the classical Newton's method for solving the *equation* $g(x,y)=0$, whose local quadratic convergence near $(\overline{x},\overline{y})$ with $\nabla g(\overline{x},\overline{y})$ nonsingular is well-known. We conclude with the following theorem, which parallels Theorem 8.

THEOREM 9. *Let $\overline{x},\overline{y},\overline{c} := c(\overline{x})$, and $\mathcal{M}_{\overline{c}}$ be as in Assumption 3. Then, there exists a neighborhood $\mathcal{N}$ of $(\overline{x},\overline{y})$ on which the conclusions of Lemma 20 are satisfied such that if $(x^0,y^0)\in\mathcal{N}$, then there exists a unique sequence $\{(x^k,y^k)\}_{k\in\mathbb{N}}$ satisfying the optimality conditions of $\mathbf{P}_k$ for all $k\in\mathbb{N}$, with*
(a) $c(x^{k-1}) + \nabla c(x^{k-1})[x^k - x^{k-1}] \in \mathcal{M}_{\overline{c}}$,
(b) $\partial h(c(x^{k-1}) + \nabla c(x^{k-1})[x^k - x^{k-1}]) = \{y^k\}$,
(c) $H_{k-1}[x^k - x^{k-1}] + \nabla c(x^{k-1})^\top y^k = 0$,
(d) $x^k - x^{k-1}$ *is a strong local minimizer of* $d \mapsto h(c(x^{k-1}) + \nabla c(x^{k-1})d) + \frac{1}{2} d^\top H_{k-1} d$.
Moreover, the sequence (x^k,y^k) converges to $(\overline{x},\overline{y})$ at a quadratic rate.

Appendix

The model function minimized in $\widehat{\mathbf{P}}$ plays a pivotal role in our analysis. Here, we establish properties of this function.

Let f be as in $\mathbf{P}$ and $\boldsymbol{u} := (\widehat{x},\widehat{y}) \in \operatorname{dom}(f)\times\mathbb{R}^m$. Define $\widehat{H} := \nabla^2_{xx}L(\widehat{x},\widehat{y})$,

$$\psi(v,w) := h(v) + w, \text{ and } \Phi_{\boldsymbol{u}}(d) := \begin{pmatrix} c(\widehat{x}) + \nabla c(\widehat{x})d \\ \frac{1}{2} d^\top \widehat{H} d \end{pmatrix}.$$

Then, for any $(v,w)\in\operatorname{dom}(h)\times\mathbb{R}$ and $(d,s)\in\mathbb{R}^n\times\mathbb{R}$,

$$\nabla\Phi_{\boldsymbol{u}}(d) = \begin{pmatrix} \nabla c(\widehat{x}) \\ d^\top \widehat{H} \end{pmatrix}, \quad \psi'((v,w);(d,s)) = h'(v;d) + s, \quad \psi''((v,w);(d,s)) = h''(v;d).$$

Define the model function $\phi_{\boldsymbol{u}}(d) := \psi(\Phi_{\boldsymbol{u}}(d)) = h(c(\widehat{x}) + \nabla c(\widehat{x})d) + \frac{1}{2}d^\top \widehat{H} d$. By [37, Theorem 13.14], $\phi_{\boldsymbol{u}}$ is piecewise linear-quadratic, though not necessarily convex because $\widehat{H}$ may not be positive semi-definite. However, $\phi_{\boldsymbol{u}}$ is convex-composite with ψ piecewise linear-quadratic convex.

The following lemma shows that if f satisfies (BCQ) at $\widehat{x}$, then the model function at $\widehat{x}$ satisfies its (BCQ) throughout its domain.

LEMMA 21. *Let f be as in $\mathbf{P}$, $\boldsymbol{u} := (\widehat{x},\widehat{y}) \in \operatorname{dom}(f)\times\mathbb{R}^m$, and suppose f satisfies (BCQ) at $\widehat{x}$. Then, $\phi_{\boldsymbol{u}}$ satisfies (BCQ) at all points $\overline{d} \in \operatorname{dom}(\phi_{\boldsymbol{u}}) = \left\{ d \;\middle|\; c(\widehat{x}) + \nabla c(\widehat{x})d \in \operatorname{dom}(h) \right\}$.*

Proof. Let $\overline{d} \in \left\{ d \,\middle|\, c(\widehat{x}) + \nabla c(\widehat{x}) d \in \operatorname{dom}(h) \right\}$. By definition,

$$\operatorname{Null}\left(\nabla\Phi_{\boldsymbol{u}}(\overline{d})^\top\right) = \operatorname{Null}\left(\begin{pmatrix}\nabla c(\widehat{x})^\top & \widehat{H} d\end{pmatrix}\right) \text{ and } N\left(\Phi_{\boldsymbol{u}}(\overline{d}) \,|\, \operatorname{dom}(\psi)\right) = N\left(c(\widehat{x}) + \nabla c(\widehat{x})\overline{d} \,|\, \operatorname{dom}(h)\right) \times \{0\}.$$

Suppose $v = (v_1, v_2) \in \operatorname{Null}\left(\nabla\Phi_{\boldsymbol{u}}(\overline{d})^\top\right) \cap N\left(\Phi_{\boldsymbol{u}}(\overline{d}) \,|\, \operatorname{dom}(\psi)\right)$. Then $v_2 = 0$, and

$$v_1 \in \operatorname{Null}\left(\nabla c(\widehat{x})^\top\right) \cap N\left(c(\widehat{x}) + \nabla c(\widehat{x})\overline{d} \,|\, \operatorname{dom}(h)\right) \subset \operatorname{Null}\left(\nabla c(\widehat{x})^\top\right) \cap N\left(c(\widehat{x}) \,|\, \operatorname{dom}(h)\right) = \{0\},$$

where the inclusion follows since $\left\langle v_1, \nabla c(\widehat{x})\overline{d} \right\rangle = 0$. □

LEMMA 22. *Let f be as in* **P**, $\boldsymbol{u} := (\widehat{x}, \widehat{y}) \in \operatorname{dom}(f) \times \mathbb{R}^m$, *and suppose f satisfies* (BCQ) *at $\widehat{x}$. Then, the cone of non-ascent directions $D_{\phi_{\boldsymbol{u}}}(\overline{d})$ at any $\overline{d} \in \operatorname{dom}(\phi_{\boldsymbol{u}})$ is given by*

$$D_{\phi_{\boldsymbol{u}}}(\overline{d}) = \left\{ \delta \,\middle|\, h'(c(\widehat{x}) + \nabla c(\widehat{x})\overline{d}; \nabla c(\widehat{x})\delta) + \overline{d}^\top \widehat{H}\delta \le 0 \right\}. \tag{77}$$

Moreover, the second-order necessary and sufficient conditions of Theorem 2 applied to $\min \phi_{\boldsymbol{u}}$ *are*

1. *If $\phi_{\boldsymbol{u}}$ has a local minimum at $\overline{d}$, then $0 \in \widehat{H}\overline{d} + \nabla c(\widehat{x})^\top \partial h(c(\widehat{x}) + \nabla c(\widehat{x})\overline{d})$ and*

$$h''(c(\widehat{x}) + \nabla c(\widehat{x})\overline{d}; \nabla c(\widehat{x})\delta) + \delta^\top \widehat{H}\delta \ge 0,$$

   for all $\delta \in D_{\phi_{\boldsymbol{u}}}(\overline{d})$.

2. *If $0 \in H\overline{d} + \nabla c(\widehat{x})^\top \partial h(c(\widehat{x}) + \nabla c(\widehat{x})\overline{d})$ and*

$$h''(c(\widehat{x}) + \nabla c(\widehat{x})\overline{d}; \nabla c(\widehat{x})\delta) + \delta^\top \widehat{H}\delta > 0,$$

   for all $\delta \in D_{\phi_{\boldsymbol{u}}}(\overline{d}) \setminus \{0\}$, then $\overline{d}$ is a strong local minimizer of $\phi_{\boldsymbol{u}}$.

Proof. Since (BCQ) is satisfied at all points $d \in \operatorname{dom}(\phi_{\boldsymbol{u}})$, the chain rule of Theorem 1 gives

$$\partial\phi_{\boldsymbol{u}}(d) = \widehat{H}d + \nabla c(\widehat{x})^\top \partial h(c(\widehat{x}) + \nabla c(\widehat{x})d),$$

$$\mathrm{d}\phi_{\boldsymbol{u}}(d)(\delta) = h'(c(\widehat{x}) + \nabla c(\widehat{x})d; \nabla c(\widehat{x})\delta) + d^\top \widehat{H}\delta,$$

which is (77). The set of Lagrange multipliers for $\phi_{\boldsymbol{u}}$ becomes

$$\begin{aligned} M_{\phi_{\boldsymbol{u}}}(d) &:= \operatorname{Null}\left(\nabla\Phi_{\boldsymbol{u}}(d)^\top\right) \cap \partial\psi(\Phi_{\boldsymbol{u}}(d)) \\ &= \operatorname{Null}\left(\begin{pmatrix}\nabla c(\widehat{x})^\top & \widehat{H}d\end{pmatrix}\right) \cap (\partial h(c(\widehat{x}) + \nabla c(\widehat{x})d) \times \{1\}), \end{aligned} \tag{78}$$

so that $\begin{pmatrix}y_1 & y_2\end{pmatrix} \in M_{\phi_{\boldsymbol{u}}}(d) \Longleftrightarrow \left\{ \widehat{H}d + \nabla c(\widehat{x})^\top y_1,\ y_1 \in \partial h(c(\widehat{x}) + \nabla c(\widehat{x})d),\ y_2 = 1. \right.$ The Lagrangian [4] is $L(d, y) := \langle y, \Phi_{\boldsymbol{u}}(d)\rangle - \psi^\star(y), \quad y = (y_1, y_2) \in \mathbb{R}^m \times \mathbb{R}$, with $\nabla^2_{dd} L(d, y) = y_2 \widehat{H}$. Then, from Theorem 2, for any $\delta \in \mathbb{R}^n$,

$$\psi''(\Phi_{\boldsymbol{u}}(d); \nabla\Phi_{\boldsymbol{u}}(d)\delta) + \max\left\{ \langle \delta, \nabla^2_{dd} L(d, y)\delta \rangle \,\middle|\, y \in M_{\phi_{\boldsymbol{u}}}(d) \right\} = h''(c(\widehat{x}) + \nabla c(\widehat{x})\overline{d}; \nabla c(\widehat{x})\delta) + \delta^\top \widehat{H}\delta.$$

□

The final lemma of this section characterizes the directions of non-ascent for the model function $\phi_{\boldsymbol{u}}$ in the presence of an active manifold. The proof is identical to Lemma 13 using $D_{\phi_{\boldsymbol{u}}}(\overline{d}) = \left\{ \delta \in \mathbb{R}^n \,\middle|\, \psi'(\Phi_{\boldsymbol{u}}(\overline{d}); \nabla\Phi_{\boldsymbol{u}}(\overline{d})\delta) \le 0 \right\}$.

LEMMA 23 **(Model non-ascent directions)**. *Let h and f be as in* **P**, *with* $\mathrm{dom}\,(h)$ *given by the Rockafellar-Wets PLQ representation of Theorem 4. Let* $(\overline{x},\overline{y}) \in \mathrm{dom}\,(f) \times \mathbb{R}^m$, *and set* $\overline{c} := c(\overline{x})$, $\overline{k} := |\mathcal{K}(\overline{c})|$, *where* $\mathcal{K}(\overline{c}) = \left\{k_1,\ldots,k_{\overline{k}}\right\}$ *are the active indices given in Definition 7. Let* $\mathcal{M}_{\overline{c}}$ *be the active manifold defined in* (34). *Let* $\boldsymbol{u} := (\widehat{x},\widehat{y}) \in \mathrm{dom}\,(f) \times \mathbb{R}^m$, *suppose f satisfies* (BCQ) *at $\widehat{x}$ and* $(\overline{d},\overline{y})$ *satisfy* $0 = \widehat{H}\overline{d} + \nabla c(\widehat{x})^\top \overline{y}$, $c(\widehat{x}) + \nabla c(\widehat{x})\overline{d} \in \mathcal{M}_{\overline{c}}$, *and* $\overline{y} \in \mathrm{ri}\left(\partial h(c(\widehat{x}) + \nabla c(\widehat{x})\overline{d})\right)$. *Then,* $\phi_{\boldsymbol{u}}$ *satisfies* (SC) *at* $\overline{d}$ *for* $(\overline{y},1)$, *and*

under Assumption 2, in the notation of Definition 12, $D_{\phi_{\boldsymbol{u}}}(\overline{d}) = \mathrm{Null}\left(A^\top \nabla c(\widehat{x})\right)$.
under Assumption 3, $D_{\phi_{\boldsymbol{u}}}(\overline{d}) = \mathbb{R}^n$.

Acknowledgments. The authors thank Asen Dontchev for helpful discussions of [8].

References

[1] Aravkin AY, Burke JV, Pillonetto G (2013) Sparse/robust estimation and Kalman smoothing with nonsmooth log-concave densities: Modeling, computation, and theory. *The Journal of Machine Learning Research* 14(1):2689–2728.

[2] Aravkin AY, Burke JV, Pillonetto G (2014) Optimization viewpoint on Kalman smoothing with applications to robust and sparse estimation. *Compressed Sensing & Sparse Filtering*, 237–280 (Springer).

[3] Burke JV (1985) Descent methods for composite nondifferentiable optimization problems. *Mathematical Programming* 33(3):260–279.

[4] Burke JV (1987) Second order necessary and sufficient conditions for convex composite NDO. *Mathematical Programming* 38(3):287–302.

[5] Burke JV, Ferris MC (1993) Weak sharp minima in mathematical programming. *SIAM Journal on Control and Optimization* 31(5):1340–1359.

[6] Burke JV, Ferris MC (1995) A Gauss-Newton method for convex composite optimization. *Mathematical Programming* 71(2):179–194.

[7] Burke JV, Poliquin R (1992) Optimality conditions for non-finite valued convex composite functions. *Mathematical Programming* 57(1):103–120.

[8] Cibulka R, Dontchev A, Kruger A (2016) Strong metric subregularity of mappings in variational analysis and optimization. *Journal of Mathematical Analysis and Applications* .

[9] Cui Y, Pang JS, Sen B (2018) Composite difference-max programs for modern statistical estimation problems. *arXiv preprint arXiv:1803.00205* .

[10] Davis D, Drusvyatskiy D (2018) Stochastic model-based minimization of weakly convex functions. *arXiv preprint arXiv:1803.06523* .

[11] Davis D, Drusvyatskiy D, Paquette C (2017) The nonsmooth landscape of phase retrieval. *arXiv preprint arXiv:1711.03247* .

[12] Deng S (1996) On uniqueness of Lagrange multipliers in composite optimization. *Journal of mathematical analysis and applications* 201(3):689–696.

[13] Dontchev AL, Rockafellar RT (2014) *Implicit functions and solution mappings* (Springer).

[14] Drusvyatskiy D, Ioffe AD, Lewis AS (2014) Clarke subgradients for directionally Lipschitzian stratifiable functions. *Mathematics of Operations Research* 40(2):328–349.

[15] Drusvyatskiy D, Lewis AS (2018) Error bounds, quadratic growth, and linear convergence of proximal methods. *Mathematics of Operations Research* .

[16] Drusvyatskiy D, Paquette C (2016) Efficiency of minimizing compositions of convex functions and smooth maps. *ArXiv e-prints* .

[17] Duchi J, Ruan F (2017) Stochastic methods for composite optimization problems. *arXiv preprint arXiv:1703.08570* .

[18] Duchi JC, Ruan F (2017) Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval. *ArXiv e-prints* .

[19] Fletcher R (2013) *Practical methods of optimization* (John Wiley & Sons).

[20] Hare W, Lewis AS (2004) Identifying active constraints via partial smoothness and prox-regularity. *Journal of Convex Analysis* 11(2):251–266.

[21] Hu Y, Li C, Yang X (2016) On convergence rates of linearized proximal algorithms for convex composite optimization with applications. *SIAM Journal on Optimization* 26(2):1207–1235.

[22] Kawasaki H (1988) Second-order necessary conditions of the Kuhn-Tucker type under new constraint qualifications. *Journal of Optimization Theory and Applications* 57(2):253–264.

[23] Lewis AS (2002) Active sets, nonsmoothness, and sensitivity. *SIAM Journal on Optimization* 13(3):702–725.

[24] Lewis AS, Wright SJ (2016) A proximal method for composite minimization. *Mathematical Programming* 158(1-2):501–546.

[25] Li C, Ng K (2007) Majorizing functions and convergence of the Gauss-Newton method for convex composite optimization. *SIAM Journal on Optimization* 18(2):613–642.

[26] Li C, Wang X (2002) On convergence of the Gauss-Newton method for convex composite optimization. *Mathematical programming* 91(2):349–356.

[27] Osborne MR (2001) *Simplicial algorithms for minimizing polyhedral functions* (Cambridge University Press).

[28] Overton ML, Ye X (1994) Towards second-order methods for structured nonsmooth optimization. *Advances in optimization and numerical analysis*, 97–109 (Springer).

[29] Powell MJ (1978) Algorithms for nonlinear constraints that use Lagrangian functions. *Mathematical programming* 14(1):224–248.

[30] Powell MJ (1978) A fast algorithm for nonlinearly constrained optimization calculations. *Numerical analysis*, 144–157 (Springer).

[31] Robinson SM (1972) A quadratically-convergent algorithm for general nonlinear programming problems. *Mathematical Programming* 3(1):145–156, ISSN 1436-4646, URL http://dx.doi.org/10.1007/BF01584986.

[32] Robinson SM (1974) Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear-programming algorithms. *Mathematical Programming* 7(1):1–16, ISSN 1436-4646, URL http://dx.doi.org/10.1007/BF01585500.

[33] Robinson SM (1980) Strongly regular generalized equations. *Mathematics of Operations Research* 5(1):43–62.

[34] Rockafellar RT (1985) Extensions of subgradient calculus with applications to optimization. *Nonlinear Analysis: Theory, Methods & Applications* 9(7):665–698.

[35] Rockafellar RT (1989) Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives. *Mathematics of Operations Research* 14(3):462–484.

[36] Rockafellar RT (2015) *Convex analysis* (Princeton University Press).

[37] Rockafellar RT, Wets RJB (1998) *Variational analysis* (Springer).

[38] Womersley RS (1985) Local properties of algorithms for minimizing nonsmooth composite functions. *Mathematical Programming* 32(1):69–89.

[39] Wright SJ (1987) Local properties of inexact methods for minimizing nonsmooth composite functions. *Mathematical programming* 37(2):232.

[40] Yuan Y (1985) On the superlinear convergence of a trust region algorithm for nonsmooth optimization. *Mathematical Programming* 31(3):269–285.